

**GRIPS Discussion Paper 14-11**

# **The Henry George Theorem in a second-best world**

**Kristian Behrens  
Yoshitsugu Kanemoto  
Yasusada Murata**

**August 2014**



**GRIPS**

NATIONAL GRADUATE INSTITUTE  
FOR POLICY STUDIES

**National Graduate Institute for Policy Studies  
7-22-1 Roppongi, Minato-ku,  
Tokyo, Japan 106-8677**

# The Henry George Theorem in a second-best world

Kristian Behrens\* Yoshitsugu Kanemoto† Yasusada Murata‡

## Abstract

The Henry George Theorem (HGT) states that, in first-best economies, the fiscal surplus of a city government that finances the Pigouvian subsidies for agglomeration externalities and the costs of local public goods by a 100% tax on land is zero at optimal city sizes. We extend the HGT to distorted economies where product differentiation and increasing returns are the sources of agglomeration economies and city governments levy property taxes. Without relying on specific functional forms, we derive a second-best HGT that relates the fiscal surplus to the excess burden expressed as an extended Harberger formula.

Keywords: Henry George Theorem; second-best economies; optimal city size; monopolistic competition; local public goods

JEL Classification: D43; R12; R13

---

\* Canada Research Chair, Département des Sciences Economiques, Université du Québec à Montréal (UQAM), Canada; National Research University, Higher School of Economics, Russia; CIRPÉE, Canada; and CEPR. E-mail: behrens.kristian@uqam.ca

† National Graduate Institute for Policy Studies (GRIPS); and Graduate School of Public Policy (GraSPP), University of Tokyo. E-mail: kanemoto@grips.ac.jp

‡ Advanced Research Institute for the Sciences and Humanities (ARISH), Nihon University, Japan. E-mail: murata.yasusada@nihon-u.ac.jp

## 1. Introduction

The equilibrium sizes of agglomerations such as cities (or communities and shopping centers), are determined by the balance between increasing and decreasing returns to spatial concentration. As is well known, cities need not be optimally sized at equilibrium since the urban environment is replete with externalities (Henderson, 1974; Kanemoto, 1980). Recent empirical research suggests that any departure from optimal city size can generate sizeable economic costs, especially in terms of foregone productivity (e.g., Au and Henderson, 2006a, 2006b). Hence, elaborating efficient urban growth policies is likely to be of first-order importance to many countries, especially developing ones. An important preliminary step for devising such urban policies is to assess whether cities are too large or too small, and by how much. The ‘golden rule’ of local public finance (Flatters et al., 1974), i.e., the *Henry George Theorem* (henceforth HGT; Stiglitz, 1977; Arnott and Stiglitz, 1979; Kanemoto, 1980; Schweizer, 1983) provides a condition for the optimal size of a city or, equivalently, the optimal number of cities given a fixed total population.<sup>1</sup> It is thus a potentially useful tool that can allow policy makers to assess whether cities are too large or too small.<sup>2</sup> Another application of the HGT is concerned with smaller agglomerations such as shopping centers and business subcenters that are mostly developed by private companies. Those developers capture increases in land prices to finance development costs. The HGT shows that free entry of developers yields an efficient allocation in a first-best world. It would be of interest for policy makers to know whether developments are too few or too many in a more realistic setting with various distortions.<sup>3</sup>

The HGT may be viewed as an extension of the result on the optimal number of firms in an industry – in the first best, entry into an industry is optimal when the marginal social benefit (i.e., the profit) of the last entrant vanishes. In a spatial context, this optimality condition must be extended to include land rents: aggregate land rents (which capitalize agglomeration benefits) equal the aggregate losses from increasing returns activities that generate agglomeration. For example, if the driving force for agglomeration is a pure local public good, then the cost of its supply must be equal to aggregate land rents at the optimal city size. In the case of a factory town, where firm-level scale economies generate spatial concentration, aggregate land rents must be equal to the losses that the firm would incur if it were constrained to price at marginal cost. In a new economic geography (henceforth NEG) model with increasing returns

---

<sup>1</sup> See Mieszkowski and Zodrow (1989) and Arnott (2004) for an overview of the HGT.

<sup>2</sup> Kanemoto et al. (1996, 2005) applied these ideas to empirically test the often-made claim that Tokyo, with a metropolitan population of about 30 million, is much too large. It is not easy to obtain reliable estimates of key variables such as the aggregate land rents and the aggregate Pigouvian subsidies in a city, but the HGT provides a promising theoretical framework for empirical studies. As stated by Arnott (2004, pp.1086–1087): “Does the Henry George Theorem provide a practical guide to optimal city size? The jury is not yet in, but the approach is sufficiently promising to merit further exploration.”

<sup>3</sup> In reality, of course, agglomeration benefits of a shopping center or a business subcenter spill over to other areas, and a developer is unlikely to capture all the benefits if he does not control all those areas. In such a case, the government may want to subsidize or share the burden of infrastructure development. Our analysis shows that, on top of the spillover effects, the government must take into account various second-best issues.

and product differentiation, aggregate land rents must be equal to the subsidies paid to firms in order to achieve efficient production scale and optimum product diversity. More generally, using the concept of the *fiscal surplus* – defined as the net surplus of the city government (property tax revenue minus costs of local public goods and any other losses from increasing returns) plus all the land rents – the HGT at the first best simply states that the fiscal surplus is zero when the number of cities is optimal (or, alternatively, when the cities are of optimal size).

As is well known, the HGT holds in a first-best world without distortions but not necessarily in a second-best world (see Arnott, 2004, for a recent survey). However, as highlighted by the latter two foregoing examples, most factors that drive the concentration of economic activity involve some form of market failure.<sup>4</sup> Thus, for the HGT to be of practical relevance it must be extended to cope with settings encompassing distortions of various kinds. This has, to the best of our knowledge, not been systematically done to date. The purpose of this article is to fill that gap by identifying conditions under which the HGT holds even in a second-best economy – an economy where policy makers can implement the optimal size of an agglomeration but are constrained to take production and consumption decisions (including entry decisions of firms) and the resulting equilibrium prices as given. We also examine in which directions the theorem needs to be modified should it fail to hold in such a world.

Several related articles have examined variations on the HGT in a second-best world. First, Arnott (2004, p.1073) showed that “in distorted urbanized economies the generalized HGT continues to hold when the aggregate magnitudes are valued at shadow prices.” The article does not, however, suggest how these shadow prices can be calculated, thereby limiting the practical relevance of the generalized HGT. Second, Helsley and Strange (1990) showed that the theorem does not hold in a matching framework of urban labor markets where firms compete for workers on a circle of skill ranges and where wages are determined via Nash bargaining. Last, Behrens and Murata (2009) examined a monocentric city model with monopolistic competition and showed that the HGT holds at the second best if and only if the second-best allocation is first-best efficient. In their model, the latter holds true only in the constant elasticity of substitution (henceforth CES) case.<sup>5</sup>

As should be clear from the foregoing literature review, whether or not the HGT holds at the second best hinges on the chosen modeling framework. Unlike previous approaches on the second-best HGT that rely from the beginning on specific functional forms, our aim is to derive a more general version of a second-best HGT and apply it to various models of spatial concentration, including those of the NEG featuring imperfect competition and product differentiation, as well as models of local public goods. Our

---

<sup>4</sup> Duranton and Puga (2004) provide an excellent survey on the theoretical foundations of agglomeration economies.

<sup>5</sup> Behrens et al. (2010) have shown that the HGT continues to hold in a CES model with heterogeneous agents and sorting along talent across cities. The reason is that, despite heterogeneity and sorting, the underlying tradeoff in terms of price and variety distortions is unaffected. Many of the results obtained in the CES model (markups, firm size, entry etc.) however are clearly knife-edge results. See Zhelobodko et al. (2012) and Dhingra and Morrow (2013) for overviews and discussions.

methodology allows us to obtain general results without resorting to specific functional forms. In a second-best world, the fiscal surplus is not generally zero at optimum and depends on two things: the *excess burden* and the *total value of distortions for residents who move from the existing cities to a new city*. Creating a new city induces changes in production and consumption in existing cities, in addition to the direct impacts of the migration of residents/workers from the existing cities to the new city. The excess burden captures the welfare effects of the induced changes, whereas the total value of distortions reflects the fact that price distortions cause divergence between the social and market values of the direct impacts. Formally, the former can be expressed as a Harberger excess burden formula (Harberger, 1964, 1971), extended to include distortions in product diversity that have largely escaped attention in the literature because no explicit price exists for the number of varieties: the weighted sum of changes in quantities and product diversity induced by the creation of a new city, with weights being the associated distortions. The latter is the weighted sum of distortions with weights being the quantities consumed by the movers.

To illustrate our results in a simple setting, we first obtain the second-best HGT in a NEG-type model where the utility function is additively separable with respect to varieties as in Dixit and Stiglitz (1977), although we follow Behrens and Murata (2007) and Zhelobodko et al. (2012) and do not restrict the functional form of the subutility a priori to the CES type. We also incorporate a property tax on housing and a pure local public good in each city. In this model, distortions in the differentiated good sector take two forms: a price distortion for each variety of the differentiated good, and a distortion for the number of varieties consumed. The variety distortion works in the opposite direction of the price distortion. We show that the price distortion depends on the relative risk aversion (henceforth RRA) and that the variety distortion is inversely related to the elasticity of the subutility function.<sup>6</sup> Unlike the RRA, the elasticity of the subutility function depends on the absolute level of utility: adding a positive constant to the subutility makes the variety distortion larger. This result reveals an important difference between models with endogenous product diversity (like new trade and NEG models) and expected utility theory that has not been emphasized much until now. In expected utility theory, utility is unique up to an affine transformation and its absolute level does not matter. In monopolistic competition models, the desirability of introducing a new variety hinges on the magnitude of the utility increase it produces. The utility increase is the difference between the utility level with equilibrium consumption and that with zero consumption. Models with endogenous product diversity, therefore, depend on the absolute level of utility, whereas expected utility theory depends only on the marginal utility and higher-order derivatives.<sup>7</sup> In our setting, whether or not the Henry George Theorem holds crucially hinges on *both* the

---

<sup>6</sup> The result that the RRA and the elasticity of the subutility matter for the gap between the first- and the second-best allocations already appears in Behrens and Murata (2009, eqs. (33) and (34)), although they do not explicitly decompose aggregate distortions into price and variety distortions.

<sup>7</sup> The fact that the variety distortion depends on the absolute level of utility is closely related to the welfare economics of new goods discussed by Romer (1994). The price of a good reflects its marginal value but does not provide the value of introducing a new good, which is given by the total utility rather than the marginal utility.

derivatives of the subutility *and* its absolute level. In other words, modeling choices related to the value of variety are important.

We then extend our framework and show that basically the same results hold in a more general model with a nonseparable utility function, general production technology, multiple differentiated goods, and congestible local public goods.

The remainder of the paper is organized as follows. In Section 2, we examine a simple model with a utility function that is additively separable with respect to varieties and a simple production technology with constant marginal and fixed costs. Section 3 extends the results to a nonseparable utility function and general production technology. Finally, Section 4 concludes.

## **2. The second-best Henry George Theorem in a simple monopolistic competition model with pure local public goods**

Before proceeding to the analysis of the general framework that encompasses many models of urban agglomeration, we illustrate the basic principles in a simple monopolistic competition model, imposing additive separability and symmetry on the demand and supply side of a differentiated good. The basic structure of the model is as follows. The economy consists of many potential sites for cities, and the number of cities that are developed is the focus of our analysis. All consumers/workers are homogeneous and free to choose a city in which to live and work. These assumptions, combined with zero moving costs, guarantee that all consumers obtain equal utility. Agglomeration economies arise from product differentiation in the consumer good, reflecting benefits from greater consumption diversity in larger cities. The size of a city is limited by the fixed supply of urban land, which causes smaller lot sizes and higher housing costs in larger cities. Housing is produced by combining land and labor, and a property tax is levied on housing. The city government provides a pure local public good.

Our problem is to obtain conditions for the optimal number of cities under the constraint that the price and variety distortions for the differentiated good cannot be removed. We solve this problem in three steps. First, assuming a fixed number of cities, we obtain the equilibrium conditions for all markets in the economy, where monopolistic competition prevails for the differentiated good. The second step is to examine the welfare impact of changing the number of cities. In this step, we use a relatively unknown version of the consumer surplus measure: the *Allais surplus* (Allais, 1943, 1977). Roughly speaking, we fix the utility level and compute the amount of the numéraire good that can be extracted from the economy. In our case, we obtain the surplus of the numéraire that is generated by an increase in the number of cities. In the final step, we exploit the fact that the equilibrium is second-best optimal if the Allais surplus is maximized there: increasing or decreasing the number of cities does not generate additional numéraire conditional on holding utility fixed at its optimal level.

### **2.1. The model**

Every potential site for cities is endowed with the same fixed amount of land,

$\bar{X}_L$ . Urban land is homogeneous and we ignore the spatial structure of a city.<sup>8</sup> Out of these potential sites,  $n$  cities are developed, where  $n$  is a potential policy variable. The total population of the economy is fixed at  $\bar{N}$ . Focusing on symmetric allocations, the size of a city is then given by  $N \equiv \bar{N}/n$ . From our assumption of free mobility and zero moving costs, all consumers achieve the same level of utility in equilibrium.

Consumers derive utility from four types of goods: leisure, a differentiated consumer good, a local public good, and housing. In addition to these consumer goods, we have another good called land that is used together with labor to produce housing. These goods cannot be transported between cities. This holds for labor as well: although workers can move between cities, once they choose a city to live in, they cannot supply labor in other cities. One special feature of land is worth emphasizing: adding a new city increases the total amount of urban land in the economy unlike other inputs such as labor. A new city does not increase the total number of workers in the economy, but it increases the total amount of land used for housing.

The utility function is  $U(x_0, U_M, x_G, x_H)$ , where  $x_0$  is the consumption of leisure,  $U_M$  is the subutility derived from the consumption of the differentiated consumer good, and  $x_G$  and  $x_H$  are the consumptions of a local public good and housing, respectively. For simplicity, we assume that the consumption of housing is fixed. This implies that an increase in the population of a city requires an increase in labor input for housing because the total available land is fixed in a city. The local public good is assumed to be purely public within a city so that all residents in a city consume the same amount of that good.

The subutility function  $U_M$  is assumed to be additively separable and symmetric with respect to varieties. Assuming that we have a continuum of varieties, it can be represented by an integral with a common lower-tier subutility function,  $u(x_j)$ :

$U_M \equiv \int_0^m u(x_j) dj$ , where  $m$  is the number (or mass) of varieties of the differentiated good. Note that this form encompasses various specifications such as the widely used CES ( $u(x) = x^{(\sigma-1)/\sigma}$ ) due to Dixit and Stiglitz (1977) and the constant absolute risk aversion (CARA) ( $u(x) = 1 - e^{-\alpha x}$ ) analyzed by Behrens and Murata (2007, 2009).<sup>9</sup>

We focus on a symmetric equilibrium where all varieties have equal prices and quantities. We also assume symmetry with respect to cities so that all cities have the same allocation. Let  $x$ ,  $p$ , and  $p_H$  respectively denote the consumption and price of a variety and the price of housing. We further assume public ownership, i.e., land and

---

<sup>8</sup> It is easy to incorporate the spatial structure within a city, for example, by using a monocentric city model as common in the first-best HGT in Arnott and Stiglitz (1979), Arnott (1979), Kanemoto (1980), and others. As shown in Behrens and Murata (2009), this extension is also possible for the second-best version.

<sup>9</sup> Note that representability by integrals does not mean additive separability. As shown in Section 3, quadratic quasi-linear preferences are not additively separable yet can be represented using integrals.

firms in all cities are owned equally by all residents.<sup>10</sup> The budget constraint can then be written as  $(\bar{x}_0 - x_0) + s = mp_x + p_H x_H$ , where  $\bar{x}_0$  and  $s$  respectively denote the endowment of time and an equal share of profits/rents minus the head tax to finance the deficit of the government.<sup>11</sup> Labor is assumed to be the numéraire.<sup>12</sup> Concerning the choice of the number of varieties to consume, a consumer is constrained by the available number of varieties in the market, denoted by  $m^S$ . A consumer then maximizes the utility function,  $U(x_0, mu(x), x_G, x_H)$ , with respect to  $x_0$ ,  $x$ , and  $m$ , subject to the budget constraint and the supply-side constraint on the mass of available varieties,  $m \leq m^S$ .

From the first-order conditions for this problem, the marginal rate of substitution between a variety and the numéraire (labor) equals their price ratio:

$$(1) \quad \frac{\partial U / \partial U_M}{\partial U / \partial x_0} u'(x) = p.$$

For the choice of the number of varieties, the first-order condition is an inequality:

$$(2) \quad \frac{\partial U / \partial U_M}{\partial U / \partial x_0} u(x) \geq px.$$

The left-hand side of (2) is the benefit from adding a variety, which should not be lower than the purchase cost of the variety. If the supply-side constraint is binding, the marginal benefit strictly exceeds the cost. Combining (1) and (2) yields the condition that the average utility is larger than or equal to the marginal utility:

$$(3) \quad \frac{u(x)}{x} \geq u'(x),$$

or, alternatively

$$(4) \quad U_R(x) \leq 1,$$

where  $U_R(x) \equiv xu'(x)/u(x)$  is the elasticity of the subutility function, which will be important when characterizing the variety distortion. We call condition (4) the ‘love-of-variety’ condition. When it does not hold, a consumer prefers to reduce the number of varieties and increase the consumption of each variety, keeping the total consumption constant. Observe also from the expression of  $U_R(x)$  that *the absolute level of utility matters* unlike in expected utility theory: shifting the utility function upward by adding for example a constant term reduces the elasticity of  $U_R(x)$  and

---

<sup>10</sup> The following analysis can be applied with minor modifications to the ‘absentee landlord case’, where we assume that rents are spent outside the economy that is modeled.

<sup>11</sup> The deficit of the government equals the cost of the local public good minus the property tax revenue.

<sup>12</sup> Strictly speaking, the numéraire is leisure in one of the cities because labor is not tradable between cities. Because of our symmetry assumption, we can take the price of leisure to be equal across cities.



tends to increase the number of varieties.

The assumption of homogeneous consumers implies that all varieties supplied in the market will be consumed because if a consumer chooses zero consumption, all other consumers will do the same, making the market demand zero. We therefore set  $m = m^S$  from now on.

The production side is formulated as follows. The differentiated consumer good and housing are produced in each city, and they cannot be transported to other cities. The sole input for the differentiated good is labor, whereas housing is produced by combining land and labor.

For the differentiated consumer good, there is only one firm producing a particular variety in a city. The number of varieties  $m$  is endogenous.<sup>13</sup> Production of a variety by a firm is denoted by  $Y$ . As noted above, production requires labor only. The fixed cost part is  $F$  and the constant marginal cost is  $c$ . We consistently adopt the notational convention that the net output is positive and the net input is negative. Labor used in producing a variety, denoted by  $Y_0^M$ , is then negative and satisfies the cost function (in labor units):  $-Y_0^M = cY + F$ . Profit maximization yields the standard condition that the price margin equals the inverse of the price elasticity of perceived demand:  $(p - c)/p = 1/\eta$ , where  $\eta$  denotes the price elasticity perceived by a producer. Under the monopolistic competition assumption with a continuum of firms, no producer has any influence on the market aggregates. In our framework, the marginal rate of substitution between the differentiated good and the numéraire in (1),  $(\partial U / \partial U_M) / (\partial U / \partial x_0)$ , is taken as fixed. This implies that the price elasticity equals the inverse of the relative risk aversion (RRA):  $\eta = 1/R_R(x)$ , where  $R_R \equiv -xu''(x)/u'(x)$  denotes the RRA, which will be important for characterizing the price distortion.<sup>14</sup> The first-order condition for profit maximization is then

$$(5) \quad \frac{p - c}{p} = R_R(x),$$

whereas the second-order condition is

$$(6) \quad -\frac{u'''(x)x}{u''(x)} \leq 2.$$

We assume that there is free entry, so that the profit of a firm is zero:  $\Pi_M = (p - c)Y - F = 0$ .

Housing is produced by combining land and labor. The housing sector is assumed to be competitive with free entry. Furthermore, all producers have the same

---

<sup>13</sup> The importance of product diversity has been emphasized by, e.g., Ethier (1982), Fujita (1989) and Fujita et al. (1999) in the new trade, urban economics, and NEG literatures.

<sup>14</sup> See Behrens and Murata (2007). Zhelobodko et al. (2012) gave a new name, the relative love of variety (RLV), to the relative risk aversion.

production technology. Under these conditions, we can represent housing production in a city by an aggregate production function with constant returns to scale:  $Y_H = F_H(Y_L^H, Y_0^H)$ , where  $Y_H$  is the total production of housing in a city and  $Y_L^H$  and  $Y_0^H$  are land and labor inputs, respectively. We assume that a property tax  $t_H$  is levied on housing. Maximizing the profit of the housing sector,  $\Pi_H = (p_H - t_H)Y_H + p_L Y_L^H + Y_0^H$ , yields the first-order conditions:

$$(7) \quad \frac{p_L}{p_H - t_H} = -\frac{\partial F_H}{\partial Y_L^H}, \quad \frac{1}{p_H - t_H} = -\frac{\partial F_H}{\partial Y_0^H}.$$

Again, from our notational convention, the inputs are negative. The maximized profit is zero from the free entry condition:  $\Pi_H = 0$ . Because housing consumption per capita is fixed, housing production in a city is determined as a function of the number of cities:  $Y_H = \bar{N}x_H / n \equiv Y_H^*(n)$ . With the land input in a city fixed at  $-Y_L^H = \bar{X}_L$ , this immediately yields the labor input required for housing production also as a function of  $n$ :  $Y_0^{H*}(n)$ .

Finally, we assume that production of the local public good requires only labor input. The latter is given by the cost function:  $-Y_0^G = C_G(Y_G)$ . Since  $Y_G$  is a pure local public good, we immediately have  $x_G = Y_G$ .

## 2.2. Equilibrium conditions

In this subsection, we examine the equilibrium conditions, given the number of cities and the supply of the local public good. Substituting the market clearing condition for a variety,  $\bar{N}x = nY$ , into the zero-profit condition yields

$$(8) \quad (p - c) \frac{\bar{N}x}{n} = F.$$

This equation, coupled with the first-order condition for profit maximization (5), determines the equilibrium price and consumption as functions of the number of cities:  $p = p^*(n)$  and  $x = x^*(n)$ . From the market clearing condition for a variety, this in turn yields the production of that variety,  $Y = Y^*(n)$ . The labor input for a variety also becomes a function of  $n$ :  $Y_0^{M*}(n) \equiv -(cY^*(n) + F)$ . Note that, given the number of cities (or equivalently the size of a city), the equilibrium price, consumption, and production of a variety are determined solely by conditions in the differentiated good sector. In particular, they are independent of the income (or utility) level of a consumer and the supply of the local public good.

Next, the number of varieties and the consumption of leisure depend on the supply of the local public good in addition to the number of cities because they must satisfy the equilibrium condition for the labor market,

$$(9) \quad \bar{N}(x_0 - \bar{x}_0) = nY_0^*(n, Y_G),$$

and the first-order conditions for utility maximization,

$$(10) \quad \frac{\partial U(x_0, mu(x^*(n)), Y_G, x_H) / \partial U_M}{\partial U(x_0, mu(x^*(n)), Y_G, x_H) / \partial x_0} u'(x^*(n)) = p^*(n),$$

where the total labor input in a city in (9) is given by:

$$(11) \quad Y_0^*(n, Y_G) \equiv mY_0^{M^*}(n) + Y_0^{H^*}(n) - C_G(Y_G).$$

Equations (9) and (10) determine the equilibrium values of leisure and variety as functions of the number of cities and the supply of the local public good:  $x_0^*(n, Y_G)$  and  $m^*(n, Y_G)$ . The utility level is then given by

$$(12) \quad U^*(n, Y_G) = U(x_0^*(n, Y_G), m^*(n, Y_G)u(x^*(n)), Y_G, x_H).$$

### 2.3. Allais surplus

Our aim is to see how the first-best HGT has to be modified in a second-best setting with monopolistic pricing, product diversity, and property tax distortions. The HGT is obtained from the first-order conditions for the optimal number of cities or, equivalently since total population is given, for the optimal size of a city. As is well known, in a first-best economy with no price distortions, indirect impacts caused by general equilibrium repercussions yield zero net benefits when measured in monetary terms. It therefore suffices to concentrate on the benefits of direct impacts. In a second-best economy, however, we also have to take into consideration the excess burden (or the deadweight loss) in addition to the direct benefits. Expressing the excess burden in pecuniary units, Harberger (1964, 1971) developed his celebrated excess burden measure, i.e., the weighted sum of induced changes in quantities consumed with the weights being the corresponding price distortions. In this paper, we take a dual approach and consider the problem of minimizing the resource cost of providing city residents a fixed level of utility.<sup>15</sup> When the resource cost is measured in terms of the numéraire good, as will be done in this paper, this leads to the consumer surplus concept developed by Allais (1943, 1977): the maximum amount of the numéraire good that can be extracted, while fixing the utility levels of all households.<sup>16</sup>

The Allais surplus provides a simple measure of welfare while being consistent. Compensating and equivalent variations (CV and EV) yield consistent consumer surplus measures, but, when applied to a general equilibrium setting, they yield much more complicated results than Harberger's. The reason is that the compensated demand

---

<sup>15</sup> Harberger in effect used the Marshallian consumer surplus to convert the utility change into pecuniary units, dividing marginal changes in utility by the marginal utility of income. This Marshallian consumer surplus has a well-known difficulty of path dependence: for a discrete change, its value is not unique and depends on the path of integration.

<sup>16</sup> Debreu (1951) proposed a variant of Allais' measure. Instead of extracting the numéraire, Debreu's coefficient of resource utilization reduces all primary inputs proportionally. Debreu's coefficient of resource utilization is not in pecuniary units, however, and as such it cannot serve as a consumer surplus measure. See Diewert (1983; 1985) and Tsuneki (1987) for further extensions.

functions do not guarantee that demand equals supply along the path of integration. This generates welfare impacts additional to direct benefits even in a first-best economy, as shown by Kanemoto and Mera (1985). Unlike the CV and EV, the Allais surplus does not have this problem because, by definition, demand equals supply in all markets where prices change.

For an arbitrary number of cities, the procedure described in the preceding subsection determines the equilibrium allocation and the utility level. Starting from this equilibrium, we examine the welfare impact of a change in the number of cities, using the Allais surplus measure. The Allais surplus is defined as the maximum amount of the numéraire good (i.e., labor) that can be extracted from the economy. Formally, we maximize

$$(13) \quad A = nY_0 - \bar{N}(x_0 - \bar{x}_0),$$

while keeping the utility level constant:

$$(14) \quad U(x_0, mu(x), Y_G, x_H) = \bar{U},$$

and all the equilibrium conditions satisfied, except, of course, for the market clearing condition for the numéraire, (9), or equivalently, the budget constraint for a consumer. In the preceding subsection, we obtained a general equilibrium with all the markets cleared, including the market for the numéraire. This yields the equilibrium utility level as a function of the number of cities. The Allais surplus uses an equilibrium in a slightly different setting where the utility level is fixed and the market clearing condition is not imposed for the numéraire. The surplus of the numéraire good in this equilibrium yields the Allais surplus. Intuitively, the equilibrium is optimal if changing the number of cities does not allow us to extract any additional surplus conditional on the equilibrium utility level.

#### 2.4. Harberger's excess burden extended to include variety distortions

As a preliminary step to obtain the optimal number of cities, we consider the effects of changing the number  $n$  of cities (or equivalently, the size  $N$  of a city given the total population  $\bar{N}$ ) on the Allais surplus:  $dA/dn$ . We assume that the supply of local public goods is fixed, relegating the analysis of the optimal supply of local public goods to Section 2.8.

As noted before, under our assumption of a separable utility function, the equilibrium consumption of a variety does not depend on the income or the utility level. The derivation of an equilibrium with a fixed utility level therefore proceeds in exactly the same way as before up to (8), and we replace the market clearing condition for the numéraire (9) by the utility constraint (14). The equilibrium values for leisure and variety are different from those in the preceding section and we denote them respectively by  $x_0^{**}(n, Y_G; \bar{U})$  and  $m^{**}(n, Y_G; \bar{U})$ . Noting (11), the Allais surplus can be written as

$$(15) \quad A = n[m^{**}(n, Y_G; \bar{U})Y_0^{M^*}(n) + Y_0^{H^*}(n) - C_G(Y_G)] - \bar{N}(x_0^{**}(n, Y_G; \bar{U}) - \bar{x}_0).$$

We assume that, given the number of cities, a unique equilibrium exists.<sup>17</sup> To ease exposition we regard  $n$  as a continuous variable.

We now examine the effect of a marginal increase in the number of cities on the Allais surplus, assuming that the supply of the local public good is fixed. Differentiating the Allais surplus with respect to the number of cities yields

$$(16) \quad \frac{dA}{dn} = Y_0 + n \left( -N \frac{dx_0}{dn} + m \frac{dY_0^M}{dn} + Y_0^M \frac{dm}{dn} + \frac{dY_0^H}{dn} \right),$$

where we omit superscripts \* and \*\* to simplify notation. This equation shows that the change in the Allais surplus consists of two parts: the first term on the right-hand side represents the change in the new city and the second term captures the impacts on the existing cities. Production in the new city requires labor input there,  $Y_0$ , but in the existing cities less labor input is required. Because the utility level is fixed, there will be no change in the surplus on the consumption side.

Applying the first-order conditions for producer and consumer optimization and the market clearing conditions for the differentiated good and housing to (16), we can decompose it into direct and indirect benefits. The former ignore general equilibrium impacts on prices and quantities whereas the latter – which corresponds to Harberger’s excess burden – captures those price and quantity effects induced by changes in the number of cities. The excess burden in our model includes the distortion for the number of varieties in addition to the price distortions. The latter are well known: the price distortion of a differentiated good is  $t \equiv p - c > 0$ , where the inequality follows from (8). The property tax creates a price distortion for housing, but this does not generate any deadweight loss because we assume that the housing consumption is fixed. In our model with an endogenous number of varieties, we must extend Harberger’s excess burden to include the distortion for the number of varieties, which we simply call the *variety distortion*. The variety distortion is the difference between the marginal social benefit and the marginal social cost of increasing the number of varieties. Because all consumers in a city benefit from the variety increase, the marginal social benefit of a variety is the sum of the marginal rates of substitution between the number of varieties and the numéraire in (2) over all residents in a city:

$$(17) \quad P_m \equiv N \frac{\partial U / \partial U_M}{\partial U / \partial x_0} u(x) = Np \frac{u(x)}{u'(x)},$$

where the second equality results from the first-order condition for utility maximization (1). The foregoing expression shows that the number of varieties has the aspect of a local public good because the benefits of an extra unit are summed over all individuals to obtain the aggregate benefit. The marginal social cost is the production cost of an additional variety,  $cY + F$ . The variety distortion is then defined as the difference between the marginal social benefit of an additional variety and the marginal social cost

---

<sup>17</sup> Though this assumption is a strong one, which need not be satisfied in general, it will be satisfied in the various applications we present in the remainder of this paper.

of that variety:  $T_m \equiv P_m - (cY + F) = [pY/u'(x)][u(x)/x - u'(x)] \geq 0$ , where we use the zero-profit condition,  $\Pi_M = pY - (cY + F) = 0$ , the market clearing condition,  $\bar{N}x = nY$ , and the love of variety condition ( 3 ) for the differentiated good. The following proposition decomposes the change in the Allais surplus into the direct benefit and Harberger's excess burden.

**Proposition 1 (Extension of Harberger's excess burden formula).** Let  $t \equiv p - c > 0$  and  $T_m \equiv P_m - (cY + F) \geq 0$  be the price distortion of a variety and the variety distortion of the differentiated good, respectively. Then, an increase in the number of cities changes the Allais surplus by

$$(18) \quad \frac{dA}{dn} = DB - EB,$$

where

$$(19) \quad DB = Y_0 + N[m(p-t)x + (p_H - t_H)x_H]$$

and

$$(20) \quad EB = -n \left( Nmt \frac{dx}{dn} + T_m \frac{dm}{dn} \right)$$

are respectively the direct benefit and the extension of Harberger's excess burden to include the variety distortion in addition to the price distortion.

**Proof:** See Appendix A1. ■

The direct benefit of creating a city involves the migration of consumers/workers from the existing cities to the new city. Because we assume the utility level to be fixed, the direct benefit appears as an increase in the Allais surplus: the amount of labor input that can be saved by creating a new city, ignoring indirect impacts through price changes. A new city increases the labor input there but existing cities need less because they have fewer residents to serve. The first term  $Y_0 < 0$  on the right-hand side of the direct benefit ( 19 ) represents the former, and the second term captures the latter. Because the decrease in the total population of existing cities equals the population of the new city,  $N$ , the labor inputs saved are  $Nm(p-t)x > 0$  and  $N(p_H - t_H)x_H > 0$ , yielding the second term in the direct benefit.

When price distortions exist, creating a new city (or, equivalently, reducing the sizes of existing cities) changes the excess burden in the existing cities in addition to generating direct benefits. As shown by Harberger (1964, 1971), the change in the excess burden can be expressed as the weighted sum of induced changes in consumption, with weights being the corresponding price distortions. The intuition is simple. An increase in the consumption of a good benefits the consumers but increases the social cost of production. The consumer and producer prices equal the marginal benefit for the consumer and the marginal cost of production, respectively, and the difference between them is the price distortion. The net social benefit of the increase in

consumption is then given by the price distortion times the change in consumption. When the number of varieties is endogenous, the Harberger formula must be extended to include the induced change in the number of varieties multiplied by the variety distortion. Note that the excess burden does not have a term with the property tax distortion because housing consumption is assumed to be fixed. For expositional simplicity, we call the marginal increase in the excess burden simply the excess burden and denote it by  $EB$ . It is not a priori clear whether the excess burden is positive or negative. We return to this point in Section 2.6 below.

## 2.5. A second-best Henry George Theorem

We have so far considered the derivative  $dA/dn$  to obtain an extension of Harberger's excess burden formula. In this subsection, we first derive the condition for the optimal number of cities by setting  $dA/dn = 0$ , which is equivalent to  $DB = EB$  from Proposition 1.<sup>18</sup> We then rewrite  $DB$  using the zero-profit conditions to establish the second-best HGT.

**Proposition 2 (Optimal number of cities).** If the number of cities is optimal, then  $dA/dn = 0$  and the direct benefit equals the excess burden:  $DB = EB$ .

**Proof:** Immediate from Proposition 1. ■

Proposition 2 provides the condition for the *optimal number* of cities. In a symmetric case where all cities have the same allocation, this is equivalent to finding the *optimal size* of a city. Flatters et al. (1974) considered the latter problem to obtain the first-best HGT for local public goods (called the 'golden rule' there). Because the number of cities satisfies  $n = \bar{N}/N$ , we can re-express our results in terms of the optimal size of a single city by maximizing the Allais surplus (15) with respect to the population of a city:  $dA/dN = 0$ . The following corollary states the condition for the optimal city size.

**Corollary (Optimal size of a single city).** If the city size is optimal at the initial equilibrium (in which  $A = 0$  by definition), then we have  $DB_N = EB_N$ , where

$$(21) \quad DB_N \equiv (\bar{x}_0 - x_0) - [m(p-t)x + (p_H - t_H)x_H]$$

and

$$(22) \quad EB_N \equiv -\left( Nmt \frac{dx}{dN} + T_m \frac{dm}{dN} \right)$$

can be interpreted as the direct benefit and the excess burden of increasing the population of a city, respectively.

**Proof:** See Appendix A2. ■

---

<sup>18</sup> This maximization of distributable surplus  $A$  is equivalent to maximizing the common utility for a fixed level of surplus under regularity conditions well known in the duality literature.

In a first-best economy, there are no price and variety distortions:  $t = T_m = 0$ . Hence, the excess burden is zero and the condition for the optimal population of a city is that the direct benefit of adding a worker is zero. This implies that the marginal product of labor equals the consumption by a worker.<sup>19</sup> In our framework, the former is labor supply multiplied by its price (i.e., one), and the latter is the sum of the social costs of the differentiated good and housing that she consumes. The social costs of differentiated good consumption must be evaluated at marginal cost rather than at the market price. In a second-best economy with price and variety distortions, the direct benefit need not be zero but must equal the excess burden created by adding a worker in a city.

Now let us derive the second-best HGT by converting the direct benefit into its dual form that includes the total land rent in a city. First, we define the *fiscal surplus* as the total land rent in a city plus the property tax revenue minus the cost of the local public good:  $FS \equiv p_L \bar{X}_L + N t_H x_H - C_G(Y_G)$ . A positive value implies that a 100% tax on land rents plus the property tax raise enough government revenue to cover the provision of the public good and to generate a surplus. Applying the zero-profit conditions in the differentiated good and housing sectors to the direct benefit ( 19 ) yields a relationship between the fiscal surplus and the direct benefit. Since the direct benefit equals the excess burden at the second best by Proposition 2, the second-best Henry George Theorem follows immediately.

**Theorem 1 (The second-best Henry George Theorem).** When the number of cities is optimal, the fiscal surplus equals the excess burden caused by increasing the number of cities plus the total value of price distortions for residents who move from the existing cities to the new city:  $FS = EB + T$ , where the fiscal surplus is the total land rent in a city plus the property tax revenue minus the cost of local public goods:  $FS \equiv p_L \bar{X}_L + N t_H x_H - C_G(Y_G)$ ; and the total value of price distortions consists of the price distortions for the differentiated good and housing:  $T \equiv (tmx + t_H x_H)N > 0$ . The price distortion for the differentiated good equals the total fixed cost,  $mF$ , and that for housing equals the property tax revenue.

**Proof:** See Appendix A3. ■

In a first-best economy with no distortions, the fiscal surplus equals the direct benefit of adding a city, where the latter is zero when the number of cities is optimal. With price distortions, this is no longer true. The total value of price distortions for those who move from the existing cities to the new city must be subtracted from the fiscal surplus:  $FS - T = DB$ . The reason is that the direct benefit, which equals the labor input saved in the existing cities minus the labor input required in the new city, is measured by the producer-side shadow price. The fiscal surplus measured by the consumer price exceeds the direct benefit by the total value of price distortions for the

---

<sup>19</sup> This condition reproduces the simple and appealing interpretation of the first-best HGT in Flatters et al. (1974, p. 102): at the optimum all wage income is devoted to private good consumption of workers and all land rents are devoted to production of the public good and private good consumption of landowners – a golden rule result. The private good consumption of landowners does not appear in our formulation because land is collectively owned by workers.



movers. In particular, the property tax revenue that is part of the fiscal surplus is completely offset by the cost of price distortions, yielding zero net surplus.<sup>20</sup>

Theorem 1 shows that the sign of the fiscal surplus at the second-best optimum depends on the signs of the total value of price distortions and the excess burden. The former is positive because the price distortions are positive in our model, but the excess burden can be positive or negative as shown in the next subsection.<sup>21</sup>

The empirical implementation of the second-best HGT would require estimates of the price distortions, in addition to the total land rent in a city that is required for the first-best version. The tax distortions are not difficult to estimate, and there have been many attempts at estimating monopolistic price distortions (see, e.g., the recent methodology developed by De Loecker and Warzinsky, 2012). Most difficult is the estimation of variety distortion because the marginal social benefit of increasing the number of varieties is not observable directly. As discussed in Kanemoto (2013c), for differentiated intermediate goods, the aggregate production functions of final goods can be used to estimate the shadow price of increasing variety. Unfortunately, this method is not applicable to differentiated consumer goods because, contrary to output levels, utility levels are not observable. One possible solution is to use variation in housing or land prices to infer the agglomeration benefits of variety. Because the utility level is equalized across cities when agents are mobile, any benefit of increased variety is capitalized into higher land prices.

## 2.6. The sign of the excess burden

Next, we examine factors that affect the sign of the excess burden. In Proposition 1, the excess burden contains only  $dx/dn$  and  $dm/dn$  after eliminating  $dx_0/dn$  by using the relationship  $dx_0/dn = -pm(dx/dn) - (P_m/N)(dm/dn)$  in Appendix A1, which is derived from the fixed utility constraint and the first-order conditions for consumer optimization.<sup>22</sup> We now eliminate  $dm/dn$  using the same relationship to obtain the following proposition.

**Proposition 3 (The sign of the excess burden).** *The excess burden is given by*

$$(23) \quad EB = -\bar{N} \left[ pm \left( \frac{t}{p} - \frac{T_m}{P_m} \right) \frac{dx}{dn} - \frac{T_m}{P_m} \frac{dx_0}{dn} \right].$$

The price margin equals  $t/p = R_R(x) > 0$ , where  $R_R(x) \equiv -xu''(x)/u'(x)$  denotes the relative risk aversion. The variety margin satisfies  $T_m/P_m = 1 - U_R(x) \geq 0$ , where

---

<sup>20</sup> Note that both the second-best and first-best HGTs do not require the supply of a local public good to be optimal.

<sup>21</sup> Note that this may not be true in general. If there is a price subsidy, for example, the price distortion can be negative.

<sup>22</sup> Note that this condition is equivalent to the fact that the expenditure function is homogeneous of degree one with respect to prices.

$U_R(x) \equiv xu'(x)/u(x)$  denotes the elasticity of the subutility function. Furthermore,  $dx/dn \geq 0$ .

**Proof:** See Appendix A4. ■

The variety distortion part of the excess burden in Proposition 1 is now rewritten in terms of the changes in consumption of the differentiated good and the numéraire. For the differentiated good, the change in consumption must be multiplied by the difference between the price and variety margins,  $t/p - T_m/P_m$ , in addition to its price and the number of varieties. The same structure applies to the numéraire, although the difference between the price and variety margins is simply  $-T_m/P_m$  because its price margin is zero.

Since both the price and variety margins are nonnegative and less than one for the differentiated good, the sign of  $t/p - T_m/P_m$  may be positive or negative. It is easy to see that both signs are indeed possible. The price margin equals the RRA,  $t/p = R_R(x)$ , and the variety margin equals one minus the elasticity of the subutility,  $T_m/P_m = 1 - U_R(x)$ .<sup>23</sup> The variety margin depends on the *level* of subutility  $u(x)$ , although the price margin does not. Adding a positive (negative) constant,  $a$ , to the subutility function,  $\tilde{u}(x) = a + u(x)$ , makes the elasticity,  $U_R(x) = xu'(x)/u(x)$ , smaller (larger) and the variety margin larger (smaller). Furthermore, the elasticity of subutility can be made arbitrarily close to zero or one by the choice of the constant. Because the RRA is not affected by the constant term, the variety margin can be made larger (smaller) than the price margin by choosing a large (small) enough constant. The excess burden can, therefore, be positive or negative. The next subsection confirms this for a special functional form that extends the CES function. It is worth noting that  $T_m/P_m - t/p$  equals the elasticity of the elasticity of subutility:

$$(24) \quad \theta_U(x) \equiv \frac{dU_R(x)}{dx} \frac{x}{U_R(x)} = \frac{T_m}{P_m} - \frac{t}{p}.$$

From the second-order condition for a differential good producer,  $dx/dn$  is positive. As noted above, for the numéraire, the price margin is zero and the difference between the price and variety margins is  $-T_m/P_m$ , which is always negative. Unfortunately, we have not been able to obtain a general result on the sign of  $dx_0/dn$ . In a special case where consumption of leisure is fixed, as assumed in many of the

---

<sup>23</sup> The variety margin is closely linked to the measure of taste for variety in Benassy (1996). In the additively separable case, his measure of taste for variety is  $\varepsilon(m) = [m\rho'(m)]/\rho(m)$ , where  $\rho(m) = mu(x)/u(mx)$ . This satisfies  $\varepsilon(m) = 1 - [mxu'(mx)]/u(mx)$ , which is similar to our  $T_m/P_m = 1 - xu'(x)/u(x)$ . The difference is that in Benassy (1996), holding  $mx$  constant implies that a change in  $m$  has no impact on  $\varepsilon(m)$ , which is not the case with our  $T_m/P_m$ . It is also equal to what Dhingra and Morrow (2013) call the ‘social markup’, which “denotes the utility from consumption of a variety net of its resource cost.”

earlier works such as Behrens and Murata (2009), it is zero. The next subsection shows that it is positive if the upper-tier utility function  $U$  is Cobb-Douglas and the lower-tier utility function  $U_M$  is CES as in Abdel-Rahman and Fujita (1990).

## 2.7. Some functional form examples

This subsection examines examples with specific functional forms, which will allow us to establish that all different cases are possible for well-behaved utility functions.<sup>24</sup>

### Example 1: A variable RRA form (extended CES)

The first example assumes that consumption of the numéraire (leisure) is fixed in addition to housing, i.e.,  $dx_0/dn=0$ . Under this assumption, only the sign of  $T_m/P_m - t/p$  matters for characterizing the sign of the excess burden since  $dx/dn \geq 0$ .

Ogaki and Zhang (2001) proposed a family of utility functions embedding varying degrees of RRA represented by a simple extension of the CES as follows:  $v(x) = (x - \gamma)^{(\sigma-1)/\sigma}$ . In our context of monopolistic competition, analyzing the case with  $\gamma > 0$  is involved because of a possible corner solution.<sup>25</sup> We thus focus on the case where  $\gamma \leq 0$ . Since the absolute level of utility matters for the sign of the excess burden, we add a constant term to  $v(x)$ :

$$(25) \quad u(x) = \begin{cases} a + (x - \gamma)^{(\sigma-1)/\sigma} & \text{for } x > 0 \\ 0 & \text{for } x = 0 \end{cases},$$

where  $\sigma > 1$ . The RRA in this case is given by  $R_R(x) = x/[\sigma(x - \gamma)]$  and its elasticity  $\theta_R(x) \equiv (dR_R(x)/dx)(x/R_R(x))$  is equal to  $\theta_R(x) = -\gamma/(x - \gamma)$ . The specification we use hence encapsulates two different cases: (i) CRRA (constant RRA) when  $\gamma = 0$ ; and (ii) IRRA (increasing RRA) when  $\gamma < 0$ . Although it does not encapsulate the DRRA (decreasing RRA) case, we now show that the sign of the excess burden can go either way depending on the value of  $a$ .

The following result summarizes how the sign of the excess burden varies with the parameter  $a$ .

**Result 1 (Variable RRA form)** In the variable RRA case with  $\sigma > 1$ , we have

$\theta_U(x) \begin{matrix} > \\ < \end{matrix} \theta_R(x)$  for  $a \begin{matrix} > \\ < \end{matrix} 0$ . Hence, when  $a = 0$ ,  $EB > 0$  in the IRRA case ( $\gamma < 0$ ), and  $EB = 0$  in the CRRA case ( $\gamma = 0$ ). If  $a$  is negative, then the excess burden may be negative even in the IRRA case.

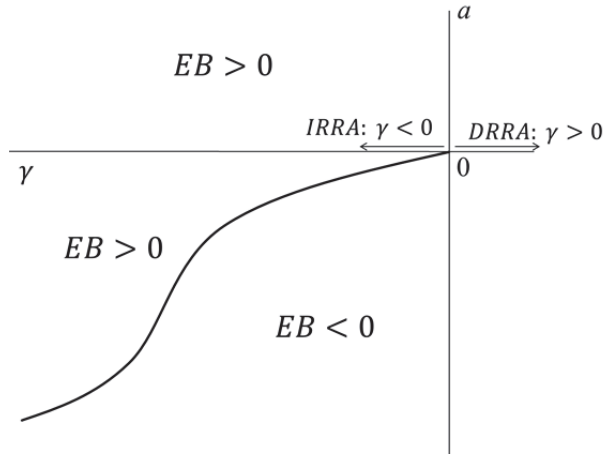
<sup>24</sup> See the discussion paper version (Behrens et al., 2010) for additional examples including CARA preferences.

<sup>25</sup> We thank Sergey Kokovin for bringing this problem to our attention.

**Proof:** See Appendix A5. ■

These results are illustrated in Figure 1. Some comments are in order. First, in the ‘standard’ CES case (with  $a = \gamma = 0$ ), the excess burden is zero, even in the second best. This confirms the results obtained by Duranton and Puga (2001) and by Behrens and Murata (2009). Second, if  $\gamma < 0$  and  $a = 0$ , the sign of the excess burden is positive. Last, an increase (decrease) in  $a$  tends to make the excess burden larger (smaller). This implies that (i) even in the CRRA case the excess burden is positive (negative) if  $a$  is positive (negative), and (ii) the excess burden can get negative even in the IRRA case. When taken together, these findings show that the case of zero excess burden is a rather special one which occurs for a zero-measure set of parameter values.

Figure 1. Variable RRA case – sign of excess burden in  $(\gamma, a)$ -space



### Example 2: Intersectoral distortions

Proposition 3 shows that even if distortions exist only in the differentiated good sector, a change in consumption of leisure induces a change in the excess burden. This is an example of intersectoral effects that are associated with changes in the variety distortion. We next examine the sign of the intersectoral effect, assuming that the consumption of the numéraire good and that of the differentiated good are aggregated via a Cobb-Douglas form and that the lower-tier utility function  $U_M$  has a CES form:

$$(26) \quad U(x_0, U_M, x_G, x_H) = U\left(x_0^\alpha \left(\int_0^m x_j^{\frac{\sigma-1}{\sigma}} dj\right)^{1-\alpha}, x_G, x_H\right),$$

where we assume that  $\sigma > 1$  and  $0 < \alpha < 1$ .<sup>26</sup> The RRA in this case is constant and

---

<sup>26</sup> This formulation is similar to that in Abdel-Rahman and Fujita (1990), but there are two major differences. First, they focus on a differentiated intermediate good, instead of a differentiated consumption good. Second, they use  $\left(\int_0^m x_j^{(\sigma-1)/\sigma} dj\right)^{\sigma/(\sigma-1)}$ , whereas we use  $\int_0^m x_j^{(\sigma-1)/\sigma} dj$  for

given by  $R_r(x) = 1/\sigma$ , so that  $t/p = 1/\sigma$ . The elasticity of utility is given by  $U_r(x) = (\sigma - 1)/\sigma$ , and hence  $T_m/P_m = 1/\sigma$ . Combining these two results yields  $\theta_U(x) = 0$ . The sign of the excess burden thus depends solely on the sign of the intersectoral distortion as captured by  $T_m/P_m$  and  $dx_0/dn$ . We can show the following result.

**Result 2 (Intersectoral distortion).** If the utility function is given by (26), then  $t/p = T_m/P_m = 1/\sigma$ , so that  $\theta_U(x) = 0$ . Furthermore,  $dx_0/dn = ((1 - \alpha)/\sigma)x_0/n$ . Hence, the excess burden is positive:

$$EB = \bar{N} \frac{T_m}{P_m} \frac{dx_0}{dn} = \frac{\bar{N}}{\sigma} \frac{dx_0}{dn} > 0$$

**Proof:** See Appendix A6. ■

Result 2 shows that, in the Cobb-Douglas/CES case, the excess burden is positive at the second-best city size. Once the intersectoral distortions disappear from the model, the excess burden is zero. This confirms Result 1: the excess burden is zero when the utility function for the differentiated good is of the CES type.

## 2.8. The optimal supply of local public goods

We now examine the optimal supply of local public goods. The Allais surplus is now a function of the local public goods in all cities in addition to the number of cities:  $A(\mathbf{Y}_G, n)$ , where  $\mathbf{Y}_G$  is a vector of the local public goods in all cities (recall that each local public good is, by definition, only consumed in the city it is provided in). Optimization with respect to each local public good  $Y_G^\ell$  requires  $\partial A / \partial Y_G^\ell = 0$  for any city  $\ell$ , where  $Y_G^\ell$  denotes the supply of the local public good in city  $\ell$ .<sup>27</sup> Because of our symmetry assumption, all cities have the same allocation at the optimum, but in order to obtain the optimality condition for a local public good we have to break symmetry and examine the effect of changing the supply of the local public good in a city with those in other cities held fixed. In the following proposition, therefore, we have to add superscript  $\ell$  to indicate city  $\ell$  in the excess burden formula.<sup>28</sup>

**Proposition 4 (The optimal supply of a local public good).** The optimality condition

---

$U_M$ . The reason for the latter choice is that we can show that as  $\alpha$  goes to zero,  $U(x_0, U_M, x_G, x_H)$  in (26) boils down to a special case of a variable RRA model that we analyze in Result 1.

<sup>27</sup> This same condition holds regardless of whether the number of cities is fixed exogenously or optimized simultaneously with the local public goods. In the latter case, the first-order conditions are simply this condition plus  $\partial A / \partial n = 0$ , and Proposition 2 remains valid as the optimality condition for the number of cities.

<sup>28</sup> Except for the excess burden, we do not have to add superscript  $\ell$ , since all variables take equal values across cities in equilibrium.

for the supply of a local public good is

$$(27) \quad P_G - C'_G(Y_G) = EB_G,$$

where

$$(28) \quad P_G = N \frac{\partial U / \partial x_G}{\partial U / \partial x_0}$$

is the marginal benefit of increasing the supply of the local public good and

$$(29) \quad EB_G = - \left( Nmt \frac{\partial x^\ell}{\partial Y_G^\ell} + T_m \frac{\partial m^\ell}{\partial Y_G^\ell} \right) - (n-1) \left( Nmt \frac{\partial x}{\partial Y_G^\ell} + T_m \frac{\partial m}{\partial Y_G^\ell} \right)$$

is the excess burden. In the partial derivatives in the excess burden formula,  $x$  and  $m$  without superscript  $\ell$  denote those in cities other than city  $\ell$ .

**Proof:** See Appendix A7. ■

Thus, the supply of the local public good is optimal when the marginal benefit minus the marginal cost equals the excess burden. The marginal benefit equals the sum of the marginal rates of substitution between the local public good and the numéraire over all city residents, and the excess burden is given by the sum of the induced changes in consumption of the differentiated good and the number of its varieties multiplied by the associated price distortions. Note that the induced changes differ between the city where the local public good is increased and the other cities.

It is worth emphasizing that equation (29) captures the interaction between the local public good, on the one hand, and the price and variety distortions, on the other hand. If there were neither price nor variety distortions (i.e.,  $t = T_m = 0$ ), then  $EB_G$  would be zero. In a second-best world, however,  $EB_G$  would not be zero, thus suggesting that ignoring the price and variety distortions may lead to an incorrect assessment of the impact of the local public good on the excess burden. This result is important, especially in an urban context, since the local public goods and product diversity (and the associated price competition) are the sources of agglomeration that have been highlighted in the classic urban economics literature and in the new economic geography, respectively.

### 3. Extension to a non-separable utility function, a general production technology, and congestible local public goods

In the additively separable cases analyzed in the preceding section, the first-order condition for profit maximization and the zero-profit condition determine  $p$  and  $x$  as functions of the number of cities,  $p^*(n)$  and  $x^*(n)$ . Hence, given  $n$ , these two variables are determined solely by the shape of the subutility function  $u(x)$ , via the relative risk aversion and the elasticity of  $u(x)$ , and the cost structure of the differentiated good sector, i.e.,  $c$  and  $F$ . This implies that the sign of  $T_m / P_m - t / p$  does

not depend on sectors other than that of the differentiated good, which allows us to obtain Proposition 3 to sign the excess burden for the special cases in Results 1 and 2. In this section we show that our main results, namely Propositions 1 – 2 and Theorem 1, remain valid without the additive separability assumption. In addition to extending the analysis to nonseparable utility functions, we assume a general production technology with multiple differentiated and other goods. We also allow for the possibility of congestion effects for local public goods.<sup>29</sup> In order to keep the analysis simple, we stick to the assumptions of identical cities and homogeneous consumers.<sup>30</sup> Furthermore, assuming that the supplies of local public goods are fixed, we concentrate on the optimal number of cities.<sup>31</sup>

### 3.1. The model

Our model has five categories of goods: the numéraire, differentiated goods, local public goods, land and other factors attached to a city, and the rest of the goods which include the structural part of housing. Let  $\mathbf{I} = \{0, 1, \dots, I\}$  denote the set of indices of goods in the economy, where good 0 is taken as the numéraire. Unlike in Section 2, the numéraire does not have to be leisure and it may or may not be tradable. The other four categories are as follows. The first category,  $\mathbf{M}$ , represents differentiated goods (either consumer goods or intermediate goods), each of which consists of differentiated varieties. We denote by  $\mathbf{m} = \{m_i\}_{i \in \mathbf{M}}$  the ‘numbers’ – or more precisely, the masses, since we work with a continuum – of varieties for each differentiated good. The second category,  $\mathbf{G}$ , contains local public goods. Both  $\mathbf{M}$  and  $\mathbf{G}$  are not tradable between cities and generate agglomeration forces. The third category,  $\mathbf{L}$ , consists of land and other factors whose supplies are *fixed* in each city at  $\bar{X}_k$ ,  $k \in \mathbf{L}$ .  $\mathbf{L}$  is also non-traded between cities but generates dispersion forces. As noted in Section 2, the most important feature of the third category is that adding a new city increases the use of these goods. We refer to them as ‘land’, but they include other resources attached to the area. The last category,  $\mathbf{H}$ , may or may not be traded, and includes all remaining goods as well as housing. For short, we call them ‘housing’. Hence  $\mathbf{I} = \{0\} \cup \mathbf{M} \cup \mathbf{G} \cup \mathbf{L} \cup \mathbf{H}$ .

The utility function of a representative consumer is given by  $U(\{U_i\}_{i \in \mathbf{M}}, \{x_i\}_{i \in \mathbf{M}})$  which consists of subutility functions  $\{U_i\}_{i \in \mathbf{M}}$  for differentiated goods and the consumption vector of other goods that include the numéraire, local public goods, land, and housing:  $\{x_i\}_{i \in \mathbf{M}} = \{x_0, \{x_i\}_{i \in \mathbf{G}}, \{x_i\}_{i \in \mathbf{L}}, \{x_i\}_{i \in \mathbf{H}}\}$ . We formulate the utility function  $U$  in a general form and impose sufficient restrictions to guarantee a unique optimum.

---

<sup>29</sup> See Berglas and Pines (1981) and Hoyt (1991) for early contributions that consider the HGT and congestible public goods. Papageorgiou and Pines (1999, Ch.10) provide a more recent treatment.

<sup>30</sup> It is not difficult to generalize the framework to heterogeneous cities and consumers, in which case the HGT holds approximately for the marginal city. See also Kanemoto (2013a) for the analysis of heterogeneous cities in the context of project evaluation.

<sup>31</sup> We do not consider the condition for the optimal supply of a public good because it is basically the same as that obtained in Section 2.8.

Each subutility function  $U_i$ , in turn, is defined over a continuum of varieties, denoted by  $\{x_{ij}, j \in [0, m_i]\}_{i \in \mathbf{M}}$ , so that  $U_i = U_i(x_{ij}, j \in [0, m_i])$ , where  $[0, m_i]$  denotes the range of varieties of differentiated good  $i$ . We focus on a class of subutility functions that allow for an integral representation and love of variety. Thus, wherever our model involves a continuum of varieties, the associated expressions are well-behaved and representable as integrals. In what follows, to alleviate notation, we denote utility by  $U(\mathbf{x}, \mathbf{m})$  for short.<sup>32</sup> We add  $\mathbf{m}$  to make it clear that utility depends on the masses of varieties via the subutility functions  $\{U_i\}_{i \in \mathbf{M}}$ . We allow for the possibility that  $U_i$  is not additively separable, like the quadratic quasi-linear utility function used by Vives (1985) and Ottaviano et al. (2002).<sup>33</sup> Finally, each consumer has an initial endowment of the numéraire,  $\bar{x}_0$ , and owns land and firms collectively. There are no endowments of differentiated goods, local public goods, and housing.

### 3.2. Consumption

Let  $\mathbf{p} = (\{p_{ij}, j \in [0, m_i]\}_{i \in \mathbf{M}}, \{p_i\}_{i \notin \mathbf{M}})$  be the vector of consumer prices, where  $\{p_{ij}, j \in [0, m_i]\}_{i \in \mathbf{M}}$  is the price vector of the differentiated goods and  $\{p_i\}_{i \notin \mathbf{M}} = \{p_0, \{p_i\}_{i \in \mathbf{G}}, \{p_i\}_{i \in \mathbf{L}}, \{p_i\}_{i \in \mathbf{H}}\}$  is that of other goods. The consumer price of a local public good here is its shadow price, i.e., the marginal rate of substitution between the local public good and the numéraire. Although local public goods are congestible, we assume that no price is charged for them. The government finances the local public goods solely by taxes and other revenues, but not by user fees.

A consumer receives income from the initial endowment of the numéraire,  $\bar{x}_0$ , as well as from the equal share,  $s$ , of land rents, profits, and the deficits of the governments:  $\bar{x}_0 + s$ . A consumer maximizes utility  $U(\mathbf{x}, \mathbf{m})$  with respect to  $\mathbf{x}$  and  $\mathbf{m}$ , subject to the budget constraint  $\bar{x}_0 + s = \sum_{i \in \mathbf{M}} \int_0^{m_i} p_{ij} x_{ij} dj + \sum_{i \notin \mathbf{M}, i \notin \mathbf{G}} p_i x_i$  and the constraints on the numbers of varieties,  $m_i \leq m_i^S$ , where  $m_i^S$  is the number of varieties available in the market for differentiated good  $i$ .

Assuming that the utility function is differentiable, we can write the first-order conditions for quantities consumed as:

---

<sup>32</sup> Note that our vector  $\mathbf{x}$  of goods may include intermediate goods, which always take the value zero for final consumers.

<sup>33</sup> The quadratic quasi-linear utility function,

$$U(\mathbf{x}, m) = \alpha \int_0^m x_j dj - \frac{\beta - \gamma}{2} \int_0^m (x_j)^2 dj - \frac{\gamma}{2} \left( \int_0^m x_j dj \right)^2 + x_0,$$

is not additively separable because it contains an interactive term as the third term on the right-hand side.



$$(30) \quad \frac{\partial U(\mathbf{x}, \mathbf{m}) / \partial x_{ij}}{\partial U(\mathbf{x}, \mathbf{m}) / \partial x_0} = p_{ij}, \quad i \in \mathbf{M}, \quad j \in [0, m_i], \quad \text{and} \quad \frac{\partial U(\mathbf{x}, \mathbf{m}) / \partial x_i}{\partial U(\mathbf{x}, \mathbf{m}) / \partial x_0} = p_i, \quad i \notin \mathbf{M}.$$

Note that the local public goods, which are not choice variables for a consumer, are included in the latter condition because we have defined their consumer (shadow) prices as the marginal rates of substitution. For the choice of the number of varieties, a consumer may have the option of not consuming all the varieties produced. However, the assumption of homogeneous consumers implies that all varieties supplied in the market will be consumed because if a consumer chooses zero consumption, all other consumers will do the same, making the market demand zero (and thus the variety would not be available in the first place). We therefore take it that a consumer consumes all varieties in the available range so that  $m_i = m_i^S$ . The number of varieties then satisfies:<sup>34</sup>

$$(31) \quad \frac{\partial U(\mathbf{x}, \mathbf{m}) / \partial m_i}{\partial U(\mathbf{x}, \mathbf{m}) / \partial x_0} \geq p_{ij} x_{ij}, \quad i \in \mathbf{M}, \quad j \in [0, m_i],$$

which is a generalization of (2). Note that  $\partial U(\mathbf{x}, \mathbf{m}) / \partial m_i$  is a shorthand for a change in utility caused by increasing the consumption of the marginal variety,  $x_{im_i}$ , from 0 to the utility-maximizing level.

### 3.3. Production

Our model has three types of products: differentiated goods, local public goods, and housing.<sup>35</sup> First, the production of differentiated good  $i \in \mathbf{M}$  in a city is represented by a transformation function  $F^i(Y_{ij}, Y_0^j, j \in [0, m_i]) = 0$ , where  $Y_{ij}$  and  $Y_0^j$  respectively denote the output of variety  $j$  and the input of the numéraire. For notational simplicity, we assume that the numéraire is the only input in the differentiated good sector.<sup>36</sup> A differentiated good firm has monopoly power and the producer shadow price diverges from the consumer price that it receives. By the free entry assumption, the profit is zero in equilibrium, however. Then, the profit of the producer of variety  $j$  of differentiated good  $i$  satisfies

$$(32) \quad \Pi_{ij} = p_{ij} Y_{ij} + Y_0^j = 0, \quad i \in \mathbf{M}, \quad j \in [0, m_i].$$

We continue to follow the sign convention that inputs take negative values and outputs take positive values.

Next, for local public goods and housing, the transformation function is

---

<sup>34</sup> For the range of varieties consumed to be represented as an interval  $[0, m_i]$ , we assume that the variety index  $j$  can be ordered so that  $(\partial U(\mathbf{x}, \mathbf{m}) / \partial m_i) / (p_{ij} x_{ij})$  evaluated at the utility-maximizing consumption level is not increasing in  $j$ . This can always be done without loss of generality.

<sup>35</sup> The numéraire can be a product but for simplicity we ignore this possibility.

<sup>36</sup> It is straightforward to introduce inputs other than numéraire.

$F^i(Y_i, Y_0, \{Y_k\}_{k \in \mathbf{L}}) = 0$ ,  $i \in \mathbf{G} \cup \mathbf{H}$ , where  $Y_i$  denotes the output of good  $i$ ; and  $Y_0$  and  $\{Y_k\}_{k \in \mathbf{L}}$  respectively denote the numéraire and land inputs. Labor and land inputs in the production of local public goods are assumed to be fixed, i.e.,  $Y_0^i = \bar{Y}_0^i$  and  $Y_k^i = \bar{Y}_k^i$  for any  $i \in \mathbf{G}$  and  $k \in \mathbf{L}$ , so that the supplies of local public goods are fixed:  $Y_i = \bar{Y}_i$  for any  $i \in \mathbf{G}$ .<sup>37</sup> However, the consumption of a local public good is endogenous because of congestion effects discussed below. The total cost of local public goods for a city is:<sup>38</sup>

$$(33) \quad C_G = -\sum_{i \in \mathbf{G}} \left( \bar{Y}_0^i + \sum_{k \in \mathbf{L}} p_k \bar{Y}_k^i \right).$$

Housing production requires the numéraire and land inputs, denoted respectively by  $Y_0^i$ ,  $i \in \mathbf{H}$ , and  $Y_k^i$ ,  $i \in \mathbf{H}$ ,  $k \in \mathbf{L}$ . As in the preceding section, we assume that the profit in the housing sector is zero. The profit in the housing sector in a city then satisfies

$$(34) \quad \Pi_i = (p_i - t_i)Y_i + Y_0^i + \sum_{k \in \mathbf{L}} p_k Y_k^i = 0, \quad i \in \mathbf{H}.$$

A property tax  $t_i$  causes price distortion in the housing sector. We assume that the property taxes are uniform across all cities.

We do not spell out the details of producer decisions here because we need a formulation that permits monopolistic as well as competitive behavior of producers. The only assumption we make is that net outputs are uniquely determined and satisfy the transformation functions.

### 3.4. Market clearing conditions

Now, we obtain the market equilibrium conditions, given the number of cities. First, market clearing for the numéraire good requires that:

$$(35) \quad \bar{N}(x_0 - \bar{x}_0) = nY_0,$$

where its total supply in a city satisfies:

$$(36) \quad Y_0 = \sum_{i \in \mathbf{M}} \int_0^{m_i} Y_0^{ij} dj + \sum_{i \in \mathbf{H}} Y_0^i + \sum_{i \in \mathbf{G}} \bar{Y}_0^i.$$

For the differentiated goods, which we assume to be non-tradable, the market clearing conditions are given by (recall that there are no initial endowments of differentiated goods):

---

<sup>37</sup> We assume that differentiated goods and housing goods are not used in the production of local public goods. Again, we could relax this assumption at the expense of heavier notation.

<sup>38</sup> Note that, unlike in Section 2 where  $C_G(Y_G)$  denotes the cost function,  $C_G$  here represents only the amount of the total cost.

$$(37) \quad Nx_{ij} = Y_{ij}, \quad i \in \mathbf{M}, \quad j \in [0, m_i].$$

As noted before, land has the special characteristic that if the number of cities increases, the total amount of land used for consumption and production increases by  $\bar{X}_i$  for  $i \in \mathbf{L}$  of the added city.<sup>39</sup> The market clearing condition for land is hence:

$$(38) \quad \bar{N}x_i - n\bar{X}_i = n \left( \sum_{k \in \mathbf{H}} Y_i^k + \sum_{k \in \mathbf{G}} \bar{Y}_i^k \right), \quad i \in \mathbf{L}.$$

For housing, which may or may not be traded, we have

$$(39) \quad \bar{N}x_i = nY_i, \quad i \in \mathbf{H}.$$

Note that the equilibrium conditions for non-traded goods must hold for each city separately, but at a symmetric equilibrium, these constraints are always satisfied. We therefore use the aggregate market clearing conditions, (35), (38), and (39), in what follows, regardless of whether the goods are traded or not.

Because migration is free and costless, the utility levels in all cities must be equalized. This equal utility condition and the above market clearing conditions, together with consumer and producer decisions, determine the equilibrium allocation. Assuming that the supply of each local public good is fixed, we focus on a class of utility and production functions such that the equilibrium is unique and write the equilibrium utility as  $U(\mathbf{x}^*(n), \mathbf{m}^*(n))$ .

### 3.5. A general version of the second-best Henry George Theorem

Assuming that the supplies of local public goods are fixed, we obtain the condition for the optimal number of cities. We do not consider the condition for the optimal supply of each public good because it is basically the same as that obtained in Section 2.8. The first step is to evaluate, in pecuniary units, the utility change driven by a change in the number of cities  $n$ . As in Section 2, we use the Allais surplus defined as the maximum amount of the numéraire good that can be extracted from the economy with the utility level being fixed at  $\bar{U} = U(\mathbf{x}^*(n), \mathbf{m}^*(n))$  obtained in the preceding subsection. For each  $n$ , given the optimal choices of consumers and producers, the market clearing conditions (37), (38), (39) determine the equilibrium values of all endogenous variables, where the market clearing condition for the numéraire, (35), is replaced by the fixed utility constraint. As said before, we assume that, given the number of cities, a unique equilibrium exists to determine the Allais surplus as a

---

<sup>39</sup> When we develop a new city, people start to use the land. In the market clearing condition, this looks like newly supplied land. Alternatively, we may think about ‘raw’ land being available in limitless supply, whereas developing a new city requires ‘constructible’ land, the supply of which increases if a new city is created. In any case, we think about land as being ‘freely available’ when developing a new city. We do not consider the conversion of non-urban land to urban land in existing cities, i.e., once a city is developed, its stock of land is fixed. We could relax this assumption and introduce also the size elasticity of city-wide surface with respect to population.

function of  $n$ :  $A(n) = nY_0(n) - \bar{N}(x_0(n) - \bar{x}_0)$ .

First, let us define price and variety distortions. Following the notation of Boadway and Bruce (1984, Chs. 8.7 and 10.3), we denote the price distortions by  $\mathbf{t} = (\{t_{ij}, j \in [0, m_i]\}_{i \in \mathbf{M}}, \{t_i\}_{i \in \mathbf{M}})$  and the shadow prices on the production side by  $\mathbf{p} - \mathbf{t}$ .<sup>40</sup> In our model, the price distortions are caused by the monopoly power in the differentiated goods markets and the property taxes. We normalize the shadow prices such that  $t_0 = 0$ .<sup>41</sup> The producer shadow prices of the differentiated goods and housing are given by the marginal rates of transformation,

$$(40) \quad p_{ij} - t_{ij} = \frac{\partial F^i / \partial Y_{ij}}{\partial F^i / \partial Y_0^{ij}}, \quad i \in \mathbf{M}, \quad j \in [0, m_i]; \quad p_i - t_i = \frac{\partial F^i / \partial Y_i}{\partial F^i / \partial Y_0^i}, \quad i \in \mathbf{H};$$

and in the housing sector the price of land satisfies

$$(41) \quad p_k = \frac{\partial F^i / \partial Y_k^i}{\partial F^i / \partial Y_0^i}, \quad i \in \mathbf{H} \quad \text{and} \quad k \in \mathbf{L},$$

where we assume differentiability of the transformation function.

As in (17), we define the consumer-side shadow price of product diversity (i.e., the value of a marginal increase in the mass of varieties  $m_i$  for good  $i \in \mathbf{M}$ ) by

$$(42) \quad P_{m_i} = N \frac{\partial U(\mathbf{x}, \mathbf{m}) / \partial m_i}{\partial U(\mathbf{x}, \mathbf{m}) / \partial x_0}, \quad i \in \mathbf{M}.$$

Note that from condition (31) the consumer-side shadow price of product diversity is larger than or equal to the total expenditure on a variety in a city:

$$(43) \quad P_{m_i} = N \frac{\partial U(\mathbf{x}, \mathbf{m}) / \partial m_i}{\partial U(\mathbf{x}, \mathbf{m}) / \partial x_0} \geq N p_{ij} x_{ij}, \quad i \in \mathbf{M}, \quad j \in [0, m_i].$$

The producer shadow price of increasing the mass of varieties of good  $i \in \mathbf{M}$  is its production cost:

$$(44) \quad P_{m_i} - T_{m_i} = -Y_0^{im_i}.$$

The local public goods require special attention. Although we are assuming that the supply of a local public good is fixed, its consumption can change when the population of a city changes, because of congestion effects. The consumption of a local public good is given by:

---

<sup>40</sup> In general, the vector of price distortions  $\mathbf{t}$  can depend on the number of cities  $n$ . We suppress this argument to alleviate the notational burden.

<sup>41</sup> As is well known in the optimal commodity tax literature, e.g., Auerbach (1985, pp.89-90), taking the price distortion to be zero on one of the goods is just a normalization.

$$(45) \quad x_i = G_i(\bar{Y}_i, N), \quad i \in \mathbf{G},$$

where we have  $\partial G_i / \partial N \leq 0$  which reflects congestion effects. The special case of a pure local public good considered in Section 2 assumes  $\partial G_i / \partial N = 0$ . Although the consumer side shadow price given by (30) is positive for a local public good, the social cost of a change in its consumption is zero because its supply and hence its production cost are fixed. For a local public good, therefore, we can take the producer shadow price to be zero, i.e.,  $p_i - t_i = 0$  for  $i \in \mathbf{G}$ . The price distortion then equals the consumer-side shadow price:  $t_i = p_i$ .

As in Section 2.4, differentiating the Allais surplus with respect to the number of cities,  $n$ , and applying the definitions of price and variety distortions yields an extension of Harberger's excess burden analogous to that from Proposition 1.

**Proposition 5 (Extension of Harberger's excess burden formula).** Let  $t_{ij}$ ,  $T_{m_i}$ , and  $t_i$  denote the price distortion of variety  $j$  of differentiated good  $i$ , the variety distortion of differentiated good  $i$ , and the price distortion of good  $i$ , respectively, where  $t_i = p_i$  for  $i \in \mathbf{G}$ . Then, an increase in the number of cities changes the Allais surplus by

$$(46) \quad \frac{dA}{dn} = DB - EB,$$

where

$$(47) \quad DB = Y_0 + N \left[ \sum_{i \in \mathbf{M}} \int_0^{m_i} (p_{ij} - t_{ij}) x_{ij} dj + \sum_{i \in \mathbf{H}} (p_i - t_i) x_i + \sum_{i \in \mathbf{L}} p_i x_i \right]$$

$$(48) \quad EB = -n \left[ N \sum_{i \in \mathbf{M}} \int_0^{m_i} t_{ij} \frac{dx_{ij}}{dn} dj + \sum_{i \in \mathbf{M}} T_{m_i} \frac{dm_i}{dn} + N \sum_{i \in \mathbf{H}} t_i \frac{dx_i}{dn} + N \sum_{i \in \mathbf{G}} t_i \frac{dx_i}{dn} \right]$$

are respectively the direct benefit and the extension of Harberger's excess burden to include variety distortions in addition to price distortions.

**Proof:** See Appendix A8. ■

Proposition 5 shows that Proposition 1 does not depend on the separability assumption and a specific production technology. The direct benefit is given by the social value of the decrease in consumption in existing cities minus the cost of supporting residents in the new city. The former must be evaluated by using the producer shadow prices when price distortions exist. The excess burden is the change in consumption multiplied by the corresponding price distortion, summed over all goods and consumers. As before, the excess burden includes variety distortions. In addition, it contains the congestion effects for local public goods, as can be seen from the last term in (48). For local public goods, we have

$$(49) \quad \frac{dx_i}{dn} = -\frac{\partial G_i}{\partial N} \frac{N}{n} \geq 0, \quad i \in \mathbf{G},$$

which is nonnegative as it is for differentiated goods. A decrease in city size caused by an increase in the number of cities reduces congestion in the consumption of the local public goods. This is captured as a decrease in the excess burden in the proposition.

Following the same procedure as in Section 2.5, we now obtain the general version of the second-best Henry George Theorem. First, if the number of cities is second-best optimal, the direct benefit equals the excess burden, yielding the generalization of Proposition 2. Next, the fiscal surplus of a local government that finances the costs of local public goods by revenues from land rents and the property tax can be defined as:

$$(50) \quad FS = \sum_{i \in \mathbf{L}} p_i \bar{X}_i + N \sum_{i \in \mathbf{H}} t_i x_i - C_G.$$

The fiscal surplus here does not contain profits because they are assumed to be zero. Applying the zero-profit conditions to the direct benefit, we can show that the fiscal surplus equals the sum of the direct benefit and the total value of price distortions for residents who move from the existing cities to the new city,  $T$ :  $FS = DB + T$ . The second-best Henry George Theorem then follows immediately.

**Theorem 2 (The second-best Henry George Theorem).** When the number of cities is optimal, the fiscal surplus equals the sum of the excess burden and the total value of price distortions for those who move from the existing cities to the new city:  $FS = EB + T$ , where the fiscal surplus is given by (50) and the total value of price distortions is:

$$(51) \quad T = \left( \sum_{i \in \mathbf{M}} \int_0^{m_i} t_{ij} x_{ij} dj + \sum_{i \in \mathbf{H}} t_i x_i \right) N.$$

**Proof:** See Appendix A9. ■

#### 4. Concluding remarks

Most factors that cause urban agglomeration involve some form of market failure. In this article, we have identified conditions under which the HGT, or the ‘golden rule’ of local public finance, holds in a second-best world and investigated in which direction the theorem needs to be modified otherwise. In so doing, we have largely focused on monopolistic competition models that are widely used in urban economics and the NEG to generate local increasing returns to scale and in which distortions in both prices and product diversity matter.

Our key findings may be summarized as follows. First, the net benefit of creating a new city can be decomposed into two parts: the direct benefit that ignores indirect impacts through price changes, and the excess burden that captures the welfare impacts of the induced changes. As shown by Harberger (1964, 1971), if price distortions exist, the excess burden can be expressed as the weighted sum of induced changes in

consumption with weights being the corresponding price distortions. In our model based on monopolistic competition, this must be extended to include the induced change in the number of varieties multiplied by the variety distortion. This result parallels those obtained by Kanemoto (2013a, b) for the benefits of transportation improvements in monopolistic competition models of urban agglomeration.

Using the zero-profit conditions, we can convert the direct benefit into its dual form that includes the total land rent in a city, thereby establishing the HGT. In a first-best economy with no distortions, the fiscal surplus equals the direct benefit, yielding the result that the fiscal surplus is zero when the number of cities is optimal. This is the ‘standard’ HGT where a 100% tax on land raises exactly the revenue required to finance the local public goods. With price distortions, this is no longer true and the total value of price distortions for residents who move from the existing cities to the new city must be subtracted from the fiscal surplus. The second-best HGT then states that the fiscal surplus equals the sum of the excess burden and the total value of price distortions, where the latter is obtained by summing, over all goods, the price distortion times the total consumption for the movers. The excess burden can be expressed as an extended Harberger Formula – the sum of induced quantity and variety changes from adding another city, weighted by the associated price and variety distortions. In the additively separable case, the sign of the excess burden is determined by the relative magnitudes of the variety and price distortions, which are in turn linked to the elasticity of the subutility and the relative risk aversion, respectively. We have shown that the price distortion tends to make the excess burden negative, while the variety distortion works in the opposite direction. When the subutility is shifted upwards, the variety distortion becomes larger and the excess burden is more likely to be positive. Last, we have shown that the HGT holds, in general, only for a zero measure subset of cases at the second best. Whether the fiscal surplus is positive or negative, i.e., whether a 100% tax on land and property tax revenue can finance the public goods at the optimal city size, depends strongly on the specification of the model.

Concerning the local public finance issues, our main results are the following three. First, the Samuelsonian condition for the optimal supply of a local public good must be extended to include the excess burden: the marginal benefit of increasing the supply of the local public good equals the marginal cost plus the excess burden. Second, if a city government levies property taxes, the fiscal surplus includes the tax revenue. However, it is also a component of the total value of price distortions, and these two offset each other. Third, a congestible local public good involves price distortion, which must be included in the excess burden. A decrease in city size reduces congestion in the consumption of the local public good, which results in a reduction in the excess burden. Last, we find that ignoring the price and variety distortions may lead to an incorrect assessment of the impact of the local public good on the excess burden. This result is important, especially in an urban context, since the local public goods and product diversity (and the associated price competition) are the sources of agglomeration that have been highlighted in the classic urban economics literature and in the new economic geography, respectively.

A first natural extension of the present work would be to depart from symmetry. As shown in Kanemoto (2013a), an extension to asymmetric cities is not difficult.

Introducing asymmetry among producers of differentiated goods, as done in the context of international trade by Melitz (2003), is more involved and left for the future. However, our result that the elasticity of the subutility and the relative risk aversion matter for the gap between the first- and the second-best allocations under firm heterogeneity, as well as for comparative static results of equilibria with respect to key variables such as market size, has been recently reconfirmed in different context (see Dhingra and Morrow, 2013, for the former, and Behrens et al., 2014, for the latter). Another direction for future work is the empirical implementation of our results following Kanemoto et al. (1996, 2005). Our more general conditions may be useful to evaluate whether observed city sizes are likely to be too large or too small. A third natural extension would be to apply our approach to the evaluation of public policies such as trade and transport policies, as they are likely to significantly affect city sizes and the spatial distribution of economic activity (Melitz, 2003; Venables, 2007). Kanemoto (2013a, b) made a first attempt in this direction for transport policies, but there are many other policy areas open for future research.

## Acknowledgments

We thank Richard Arnott, Tom Holmes, Sergey Kokovin, two anonymous referees and the handling editor, Vernon Henderson, as well as participants at the Nagoya Workshop for Spatial Economics and International Trade and at the 56th Meetings of the RSAI in San Francisco for valuable comments and suggestions. Behrens is holder of the Canada Research Chair in Regional Impacts of Globalization. Financial support from the CRC Program of the Social Sciences and Humanities Research Council (SSHRC) of Canada is gratefully acknowledged. Behrens further gratefully acknowledges financial support from FQRSC Québec (Grant NP-127178). Kanemoto gratefully acknowledges financial support from kakenhi (18330044), (23330076). Murata gratefully acknowledges financial support from kakenhi (17730165) and mex.t.academic frontier (2006-2010). Any remaining errors are ours.

## Appendix A: Proofs

### A1. Proof of Proposition 1

The proof proceeds in two steps. The first step is to convert the changes in labor inputs in the existing cities into those in outputs and consumption there, noting that they must satisfy the production and utility functions and the first-order conditions for optimal choices of producers and consumers. The second step exploits the fact that creating an additional city reduces the need for consumption goods in the existing cities. Using the market clearing conditions for the differentiated good and housing, we obtain the relationship between the changes in outputs in the existing cities and those in the new city.

First, differentiating the cost function of the differentiated good yields  $-dY_0^M / dn = c(dY / dn)$ , i.e., the change in the amount of input (with minus sign because of our sign convention) equals the change in the output multiplied by the marginal cost parameter,  $c$ . Differentiating the housing production function, applying the first-order conditions for profit maximization, and noting that the supply of land is



fixed,  $dY_L^H / dn = 0$ , we obtain:

$$(52) \quad \frac{dY_0^H}{dn} = -(p_H - t_H) \frac{dY_H}{dn}.$$

On the consumption side, differentiating the fixed utility constraint ( 14 ) and using the first-order condition for optimal consumption choice yields:

$$(53) \quad \frac{dx_0}{dn} = -pm \frac{dx}{dn} - \frac{P_m}{N} \frac{dm}{dn}.$$

Substituting these relationships into ( 16 ), we can rewrite the change in Allais surplus as:

$$(54) \quad \frac{dA}{dn} = Y_0 + n \left( Npm \frac{dx}{dn} - mc \frac{dY}{dn} + (P_m + Y_0^M) \frac{dm}{dn} - (p_H - t_H) \frac{dY_H}{dn} \right).$$

In the second step, we relate the changes in the outputs and consumption in the existing cities to those in the new city, using the market clearing conditions. For the differentiated good, differentiating the market clearing condition,  $\bar{N}x = nY$ , summing over all varieties, and applying  $Y = Nx$ , yields:

$$(55) \quad nm \frac{dY}{dn} = -m \left( Nx - \bar{N} \frac{dx}{dn} \right).$$

This shows that the reduction in the total output of the differentiated good in existing cities equals the output in one city (i.e., the additional city) minus the increase in the total consumption. Thus, the decrease in the total labor input for the differentiated good in existing cities has two components: the shift of production to the new city and the induced change in consumption caused by general equilibrium repercussions. Both have to be multiplied by the marginal cost of production. Because consumption of housing per person is fixed, differentiating the market clearing condition for housing,  $\bar{N}x_H = nY_H$ , with respect to  $n$  yields

$$(56) \quad n \frac{dY_H}{dn} = -Nx_H.$$

This shows that decreases in housing production in the existing cities exactly offset housing produced in the new city.

Substituting ( 55 ), ( 56 ), and  $Y_0^M = -(cY + F)$  into ( 54 ) yields

$$(57) \quad \frac{dA}{dn} = Y_0 + N \left[ mcx + (p_H - t_H)x_H \right] + n \left[ Nm(p - c) \frac{dx}{dn} + (P_m - (cY + F)) \frac{dm}{dn} \right].$$

The first two terms on the right-hand side represent the direct benefit and the last term the excess burden. Using the price and variety distortions for the differentiated good, we can rewrite ( 57 ) to obtain the extension of Harberger's excess burden to include the

variety distortion.

## A2. Proof of Corollary

From  $n = \bar{N} / N$ , we have  $dn / dN = -\bar{N} / N^2$ , which yields

$$(58) \quad \frac{dA}{dN} = \frac{dA}{dn} \frac{dn}{dN} = -\frac{\bar{N}}{N^2} \frac{dA}{dn} = -\frac{\bar{N}}{N^2} (DB - EB).$$

Because this is for the entire economy, we have to divide this by the number of cities to obtain the condition that corresponds to Flatters et al. (1974):

$$(59) \quad \frac{N}{\bar{N}} \frac{dA}{dN} = DB_N - EB_N,$$

where  $DB_N \equiv -DB / N$  and  $EB_N \equiv -EB / N$  can be interpreted as the direct benefit and the excess burden of increasing the population of a city. From the market clearing conditions for a variety and housing, we can express the direct benefit using the consumption side variables as:

$$(60) \quad DB_N = (\bar{x}_0 - x_0) - mcx - (p_H - t_H)x_H - \frac{A}{\bar{N}}.$$

The excess burden can be written as:

$$(61) \quad EB_N = -\left[ Nmt \frac{dx}{dN} + T_m \frac{dm}{dN} \right].$$

If the initial allocation, where  $A = 0$ , has the optimal city size, then  $dA / dN = 0$ , which immediately yields the corollary.

## A3. Proof of Theorem 1

First, Proposition 1 shows that the direct benefit in the housing sector is the before-tax value of housing minus the labor cost:  $(p_H - t_H)Y_H + Y_0^H$ . Because the profit is zero in the housing sector, this must equal the total land rent in a city:  $(p_H - t_H)Y_H + Y_0^H = p_L \bar{X}_L$ . Second, the direct benefit in the local public good sector is negative and equals its cost:  $Y_0^G = -C_G(Y_G)$ . Third, the direct benefit in the differentiated good sector is the variable cost minus the total cost in a city:  $mcY + mY_0^M$ . From  $-Y_0^M = cY + F$  and the zero-profit condition,  $pY = cY + F$ , we obtain  $mcY + mY_0^M = -tmY$ . Summing the three parts of the direct benefit then yields  $DB = p_L \bar{X}_L - C_G(Y_G) - tmY$ . This can be rewritten as  $DB = FS - T$ , where  $T \equiv tmY + t_H Y_H$ . From the zero-profit condition,  $pY = cY + F$ , the fixed cost satisfies  $F = tY$ .

## A4. Proof of Proposition 3

First, solving (53) for  $dm / dn$  and substituting the result into the excess

burden in Proposition 1 yields ( 23 ). Second, substituting ( 17 ) into the definition of the variety distortion immediately yields  $T_m / P_m = 1 - U_R(x) \geq 0$ .

Next, the sign of  $dx/dn$  can be obtained as follows. From the zero-profit condition and from  $p = c/(1 - R_R(x))$ , we readily obtain

$$\pi = (p - c)Nx - F = \frac{R_R(x)}{1 - R_R(x)} cx \frac{\bar{N}}{n} - F = 0$$

For convenience, let us rewrite this expression as follows:

$$( 62 ) \quad xR_R(x) + \phi n[R_R(x) - 1] = 0,$$

where  $\phi \equiv F/c\bar{N} > 0$  is a bundle of parameters. Differentiating this equation with respect to  $x$  and  $n$ , and using the definition of the relative risk aversion, we obtain

$$\frac{dx}{dn} = \frac{R_R(x)}{(1 - R_R(x))^2} \left( 2 + \frac{u'''x}{u''} \right) \geq 0,$$

where the inequality follows from the second-order condition for profit maximization of a differentiated good producer ( 6 ).

#### A5. Proof of Result 1

The elasticity of subutility satisfies

$$( 63 ) \quad U_R(x) = \frac{\sigma - 1}{\sigma} \frac{x}{x - \gamma} \frac{1}{1 + a(x - \gamma)^{(1-\sigma)/\sigma}},$$

which is decreasing in  $a$ . The elasticity of the elasticity of subutility is then given by

$$( 64 ) \quad \theta_U(x) = -\frac{\gamma}{x - \gamma} + \frac{\sigma - 1}{\sigma} \frac{x}{x - \gamma} \frac{a(x - \gamma)^{\frac{1-\sigma}{\sigma}}}{1 + a(x - \gamma)^{\frac{1-\sigma}{\sigma}}}.$$

Comparing this expression with  $\theta_R(x)$  and making use of the love of variety condition ( 3 ), we readily see that

$$( 65 ) \quad \theta_U(x) \begin{matrix} > \\ < \end{matrix} \theta_R(x) \text{ as } a \begin{matrix} > \\ < \end{matrix} 0.$$

In particular, if  $a = 0$ , then  $\theta_U(x)$  coincides with the elasticity of RRA and we have

(i)  $\theta_U(x) > 0$  in the IRRRA case where  $\gamma < 0$ , and (ii)  $\theta_U(x) = 0$  in the CRRA case where  $\gamma = 0$ . If  $a$  is negative, then  $\theta_U(x)$  may be negative even in the IRRRA case.

Finally, noting that  $dx/dn \geq 0$  from the proof of Proposition 3, we obtain Result 1.

## A6. Proof of Result 2

The only step that has not been proved is the sign of  $dx_0/dn$ . In order to sign this derivative, observe that the first-order condition for utility minimization ( 10 ) implies that

$$(66) \quad \frac{1-\alpha}{\alpha} \frac{\sigma-1}{\sigma} \frac{x_0}{mx} = p.$$

From  $R_R = 1/\sigma$ , the profit maximization condition ( 5 ) and the zero-profit condition ( 8 ) yields

$$(67) \quad x = (\sigma-1)\phi n \quad \text{and} \quad p = \frac{\sigma}{\sigma-1} c,$$

where  $\phi \equiv F/(c\bar{N})$  is a constant term. Plugging these into ( 66 ) above then yields a relationship between  $x_0$  and  $n$  as:

$$(68) \quad m = \frac{1-\alpha}{\alpha} \frac{\sigma-1}{\sigma^2} \frac{1}{\phi c} \frac{x_0}{n}.$$

Substituting ( 67 ) and ( 68 ) into the utility function ( 26 ) and noting the constant utility constraint then yields:

$$U \left( x_0 \left( \frac{1-\alpha}{\alpha} \frac{\sigma-1}{\sigma^2} \frac{1}{\phi c} (\sigma-1) \frac{(\sigma-1)}{\sigma} \phi \frac{(\sigma-1)}{\sigma} n \frac{1}{\sigma} \right)^{1-\alpha}, x_G, x_H \right) = \bar{U}$$

Since  $x_G$  and  $x_H$  are constant, differentiating this equation yields

$$\frac{dx_0}{dn} = \frac{x_0}{\sigma n} (1-\alpha) > 0$$

## A7. Proof of Proposition 4

From the definition of the Allais surplus, we first obtain

$$(69) \quad \begin{aligned} \frac{\partial A}{\partial Y_G^\ell} = & -C'_G(Y_G^\ell) - N \frac{\partial x_0^\ell}{\partial Y_G^\ell} + m \frac{\partial Y_0^{M\ell}}{\partial Y_G^\ell} + Y_0^M \frac{\partial m^\ell}{\partial Y_G^\ell} + \frac{\partial Y_0^{H\ell}}{\partial Y_G^\ell} + (\bar{x}_0 - x_0) \frac{\partial N^\ell}{\partial Y_G^\ell} \\ & + (n-1) \left( -N \frac{\partial x_0}{\partial Y_G^\ell} + m \frac{\partial Y_0^M}{\partial Y_G^\ell} + Y_0^M \frac{\partial m}{\partial Y_G^\ell} + \frac{\partial Y_0^H}{\partial Y_G^\ell} + (\bar{x}_0 - x_0) \frac{\partial N}{\partial Y_G^\ell} \right) \end{aligned}$$

Here, variables without superscript  $\ell$  denote those in cities other than  $\ell$ . Because we start from a symmetric equilibrium, all the variables except for the partial derivatives are the same between city  $\ell$  and other cities. We therefore drop superscript  $\ell$  for those variables.

Because the total population is fixed, we have

$$(70) \quad (\bar{x}_0 - x_0) \left[ \frac{\partial N^\ell}{\partial Y_G^\ell} + (n-1) \frac{\partial N}{\partial Y_G^\ell} \right] = 0.$$

The partial derivatives of the city population with respect to the local public good therefore drop out from ( 69 ). As for housing, production of housing is proportional to the city population:  $Y_H = \bar{x}_H N$ . From the production function and the first-order conditions for profit maximization, ( 7 ), the labor input for housing then satisfies:

$$(71) \quad \frac{\partial Y_0^{H\ell}}{\partial Y_G^\ell} = -(p_H - t_H) \frac{\partial Y_H^\ell}{\partial Y_G^\ell} = -\bar{x}_H (p_H - t_H) \frac{\partial N^\ell}{\partial Y_G^\ell}, \text{ for any city } \ell'.$$

Hence, the partial derivatives of the labor input for housing production also cancel each other out:

$$(72) \quad \frac{\partial Y_0^{H\ell}}{\partial Y_G^\ell} + (n-1) \frac{\partial Y_0^H}{\partial Y_G^\ell} = -\bar{x}_H (p_H - t_H) \left[ \frac{\partial N^\ell}{\partial Y_G^\ell} + (n-1) \frac{\partial N}{\partial Y_G^\ell} \right] = 0.$$

Concerning differentiated good production, from the cost function and the market clearing condition, we have

$$(73) \quad \frac{\partial Y_0^{M\ell}}{\partial Y_G^\ell} = -c \frac{\partial Y^\ell}{\partial Y_G^\ell} = -c \left( N \frac{\partial x^\ell}{\partial Y_G^\ell} + x \frac{\partial N^\ell}{\partial Y_G^\ell} \right), \text{ for any city } \ell'.$$

As before, the population change terms cancel each other out and we are left with

$$(74) \quad \frac{\partial Y_0^{M\ell}}{\partial Y_G^\ell} + (n-1) \frac{\partial Y_0^M}{\partial Y_G^\ell} = -cN \left[ \frac{\partial x^\ell}{\partial Y_G^\ell} + (n-1) \frac{\partial x}{\partial Y_G^\ell} \right].$$

Differentiating the constant utility condition ( 14 ) and substituting the definition of the marginal benefit of increasing the supply of the local public good in ( 28 ) yields

$$(75) \quad \frac{\partial x_0^\ell}{\partial Y_G^\ell} = -pm \frac{\partial x^\ell}{\partial Y_G^\ell} - \frac{P_m}{N} \frac{\partial m^\ell}{\partial Y_G^\ell} - \frac{P_G}{N}.$$

In other cities, we have

$$(76) \quad \frac{\partial x_0}{\partial Y_G^\ell} = -pm \frac{\partial x}{\partial Y_G^\ell} - \frac{P_m}{N} \frac{\partial m}{\partial Y_G^\ell}.$$

Substituting ( 72 ), ( 74 ), ( 75 ), and ( 76 ) into ( 69 ) yields

$$(77) \quad \begin{aligned} \frac{\partial A}{\partial Y_G^\ell} = & -C'_G(Y_G) + P_G + \left( Nm(p-c) \frac{\partial x^\ell}{\partial Y_G^\ell} + (P_m - (cY + F)) \frac{\partial m^\ell}{\partial Y_G^\ell} \right) \\ & + (n-1) \left( Nm(p-c) \frac{\partial x}{\partial Y_G^\ell} + (P_m - (cY + F)) \frac{\partial m}{\partial Y_G^\ell} \right). \end{aligned}$$

The optimality condition for the supply of the local public good immediately follows from  $\partial A / \partial Y_G^\ell = 0$ .

#### A8. Proof of Proposition 5

First, differentiating the Allais surplus  $A(n) = nY_0(n) - \bar{N}(x_0(n) - \bar{x}_0)$  with respect to the number of cities yields:

$$(78) \quad \frac{dA}{dn} = Y_0 - \bar{N} \frac{dx_0}{dn} + n \frac{dY_0}{dn}.$$

We modify this expression by using the conditions that the equilibrium values satisfy the transformation function and the fixed utility constraint. Differentiating the transformation functions,  $F^i(Y_{ij}, Y_0^{ij}, j \in [0, m_i]) = 0$  of differentiated good  $i \in \mathbf{M}$  and  $F^i(Y_i, Y_0^i, \{Y_k^i\}_{k \in \mathbf{L}}) = 0$  for housing  $i \in \mathbf{H}$ , with respect to  $n$  yields

$$(79) \quad \frac{\partial F^i}{\partial Y_{ij}} \frac{dY_{ij}}{dn} + \frac{\partial F^i}{\partial Y_0^{ij}} \frac{dY_0^{ij}}{dn} = 0 \quad \text{for } j \in [0, m_i], \quad i \in \mathbf{M},$$

$$(80) \quad \frac{\partial F^i}{\partial Y_i} \frac{dY_i}{dn} + \frac{\partial F^i}{\partial Y_0^i} \frac{dY_0^i}{dn} + \sum_{k \in \mathbf{L}} \frac{\partial F^i}{\partial Y_k^i} \frac{dY_k^i}{dn} = 0 \quad \text{for } i \in \mathbf{H}.$$

Differentiating (36) with respect to  $n$ , we obtain

$$(81) \quad \frac{dY_0}{dn} = \sum_{i \in \mathbf{M}} \left( \int_0^{m_i} \frac{dY_0^{ij}}{dn} dj + Y_0^{im_i} \frac{dm_i}{dn} \right) + \sum_{i \in \mathbf{H}} \frac{dY_0^i}{dn}.$$

Substituting (79) and (80) into this relationship and applying the definitions of the producer shadow prices, (40), (41), and (44), yields

$$(82) \quad \frac{dY_0}{dn} = - \sum_{i \in \mathbf{H}} (p_i - t_i) \frac{dY_i}{dn} - \sum_{k \in \mathbf{L}, i \in \mathbf{H}} p_k \frac{dY_k^i}{dn} - \sum_{i \in \mathbf{M}} \left( \int_0^{m_i} (p_{ij} - t_{ij}) \frac{dY_{ij}}{dn} dj + (P_{m_i} - T_{m_i}) \frac{dm_i}{dn} \right),$$

where we have used the normalizations,  $p_0 = 1$  and  $t_0 = 1$ , and the assumption that the supply of a local public good is fixed. Next, differentiating the fixed utility constraint  $\bar{U} = U(\mathbf{x}(n), \mathbf{m}(n))$  and using the first-order condition for optimal consumption choice (30), as well as the definition of the consumer-side shadow price of product diversity (42), yields:

$$(83) \quad \frac{dx_0}{dn} = - \sum_{i \neq 0, i \in \mathbf{M}} p_i \frac{dx_i}{dn} - \sum_{i \in \mathbf{M}} \left( \int_0^{m_i} p_{ij} \frac{dx_{ij}}{dn} dj + \frac{1}{N} P_{m_i} \frac{dm_i}{dn} \right).$$

Substituting these relationships into (78), we can rewrite the change in Allais surplus as:

$$(84) \quad \frac{dA}{dn} = Y_0 + \bar{N} \left[ \sum_{i \neq 0, i \in \mathbf{M}} p_i \frac{dx_i}{dn} + \sum_{i \in \mathbf{M}} \left( \int_0^{m_i} p_{ij} \frac{dx_{ij}}{dn} dj + \frac{1}{N} P_{m_i} \frac{dm_i}{dn} \right) \right] \\ - n \left[ \sum_{i \in \mathbf{H}} (p_i - t_i) \frac{dY_i}{dn} + \sum_{i \in \mathbf{L}, k \in \mathbf{H}} p_i \frac{dY_i^k}{dn} + \sum_{i \in \mathbf{M}} \left( \int_0^{m_i} (p_{ij} - t_{ij}) \frac{dY_{ij}}{dn} dj + (P_{m_i} - T_{m_i}) \frac{dm_i}{dn} \right) \right].$$

In the second step, we relate the changes in the outputs and consumption in the existing cities to those in the new city, using the market clearing conditions. Differentiating the market clearing conditions (37) – (39) with respect to  $n$  yields:

$$(85) \quad \bar{N} \frac{dx_{ij}}{dn} = Nx_{ij} + n \frac{dY_{ij}}{dn}, \quad \forall i \in \mathbf{M}, \quad \forall j \in [0, m_i]$$

$$(86) \quad \bar{N} \frac{dx_i}{dn} = Nx_i + n \sum_{k \in \mathbf{H}} \frac{dY_i^k}{dn}, \quad \forall i \in \mathbf{L}$$

$$(87) \quad \bar{N} \frac{dx_i}{dn} = Nx_i + n \frac{dY_i}{dn}, \quad \forall i \in \mathbf{H},$$

where we are assuming that the supplies of the local public goods are fixed. Substituting these into (84) yields (46), (47), and (48) in Proposition 5.

#### A9. Proof of Theorem 2

Substituting the market clearing conditions (37), (38), and (39) into the direct benefit yields:

$$(88) \quad DB = Y_0 + \left[ \sum_{i \in \mathbf{M}} \int_0^{m_i} (p_{ij} - t_{ij}) Y_{ij} dj + \sum_{i \in \mathbf{H}} (p_i - t_i) Y_i \right] + \sum_{i \in \mathbf{L}} p_i (\bar{X}_i + \sum_{k \in \mathbf{H}} Y_i^k + \sum_{k \in \mathbf{G}} \bar{Y}_i^k).$$

Using the zero-profit conditions, (32) and (34), and the definition of the total cost of local public goods, (33), we can further rewrite the direct benefit as:

$$(89) \quad DB = \sum_{i \in \mathbf{L}} p_i \bar{X}_i - C_G - \sum_{i \in \mathbf{M}} \int_0^{m_i} t_{ij} Y_{ij} dj,$$

which yields  $DB = FS - T$ .

#### References

- Abdel-Rahman, H., Fujita M., 1990. Product variety, Marshallian externalities, and city sizes. *Journal of Regional Science* 30, 165–184.
- Allais, M., 1943. *A la Recherche d'une Discipline Économique*, vol. I. Paris: Imprimerie Nationale.
- Allais, M., 1977. *Theories of General Economic Equilibrium and Maximum Efficiency*. In: Schwödiauer, G. (Ed.), *Equilibrium and Disequilibrium in Economic Theory*. D. Reidel Publishing Company, Dordrecht, pp. 129–201.

- Arnott, R., 2004. Does the Henry George Theorem provide a practical guide to optimal city size? *The American Journal of Economics and Sociology* 63, 1057–1090.
- Arnott, R., Stiglitz, J.E., 1979. Aggregate land rents, expenditure on public goods, and optimal city size. *Quarterly Journal of Economics* 93, 471–500.
- Au, C.-C., Henderson, J.V., 2006a. Are Chinese cities too small? *Review of Economic Studies* 73, 549–576.
- Au, C.-C., Henderson, J.V., 2006b. How migration restrictions limit agglomeration and productivity in China. *Journal of Development Economics* 80, 350–388.
- Auerbach, A.J., 1985. The theory of excess burden and optimal taxation. In: Auerbach, A.J., Feldstein, M. (Eds.), *Handbook of Public Economics*, vol. 1. Elsevier, Amsterdam, 61–127.
- Behrens, K., Duranton G., Robert-Nicoud, F.L., 2010. Productive cities: Sorting, selection and agglomeration. CEPR Discussion Papers 7922, C.E.P.R. Discussion Papers.
- Behrens, K., Kanemoto, Y., Murata Y., 2010. The Henry George Theorem in a second-best world. CEPR Discussion Papers 8120, C.E.P.R. Discussion Papers.
- Behrens, K., Murata, Y., 2007. General equilibrium models of monopolistic competition: A new approach. *Journal of Economic Theory* 136, 776–787.
- Behrens, K., Murata, Y., 2009. City size and the Henry George Theorem under monopolistic competition. *Journal of Urban Economics* 65, 228–235.
- Behrens, K., Pokrovsky, D., Zhelobodko, E., 2014. Market size, entrepreneurship, and income inequality. CEPR Discussion Papers 9831, C.E.P.R. Discussion Papers.
- Benassy, J.-P., 1996. Taste for variety and optimum production patterns in monopolistic competition. *Economics Letters* 52, 41–47.
- Berglas, E., Pines, D., 1981. Clubs, local public goods and transportation models: A synthesis. *Journal of Public Economics* 15, 141–162.
- Boadway, R., Bruce, N., 1984. *Welfare Economics*. Oxford: Basil-Blackwell.
- Debreu, G. 1951. The coefficient of resource utilization. *Econometrica*, 19, 273-292.
- De Loecker, J., Warzynski, F., 2012. Markups and firm-level export status. *American Economic Review*, 102(6): 2437-2471.
- Dhingra S., Morrow, J., 2013. Monopolistic competition and optimum product diversity under firm heterogeneity. LSE, mimeo.
- Diewert, W. E., 1983. The Measurement of waste within the production sector of an open economy. *Scandinavian Journal of Economics*, 85, 159-179.
- Diewert, W. E., 1985. Measurement of waste and welfare. in applied general equilibrium models. In: Piggot, J., Whalley, J. (Eds.), *New Developments in Applied General Equilibrium Analysis*. Cambridge University Press, Cambridge, New York, and Melbourne, pp. 42-102.
- Dixit, A.K., Stiglitz, J.E., 1977. Monopolistic competition and optimum product



- diversity. *American Economic Review* 67, 297–308.
- Duranton, G., Puga, D., 2001. Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review* 91, 1454–1477.
- Duranton, G., Puga, D., 2004. Micro-foundation of urban agglomeration economies. In: Henderson, J.V., Thisse, J.F. (Eds.), *Handbook of Regional and Urban Economics*, vol. 4. Elsevier, Amsterdam, 2063–2117.
- Ethier, W.J., 1982. National and international returns to scale in the modern theory of international trade. *American Economic Review* 72, 389–405.
- Flatters, F., Henderson, J.V., Mieszkowski, P., 1974. Public goods, efficiency, and regional fiscal equalization. *Journal of Public Economics* 3, 99–112.
- Fujita, M., 1989. *Urban Economic Theory: Land Use and City Size*. Cambridge Univ. Press, Cambridge, MA.
- Fujita, M., Krugman, P., Venables, A., 1999. *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, Massachusetts.
- Harberger, A.C., 1964. The measurement of waste. *American Economic Review* 54, 58–76.
- Harberger, A.C., 1971. Three basic postulates for applied welfare economics: An interpretive essay. *Journal of Economic Literature* 9, 785–797.
- Helsley, R. W., Strange, W. C., 1990. Matching and agglomeration economies in a system of cities. *Regional Science and Urban Economics* 20, 189–212.
- Henderson, J.V., 1974. The sizes and types of cities. *American Economic Review* 64, 640 – 656.
- Hoyt, W.H., 1991. Competitive jurisdictions, congestion, and the Henry George Theorem: When should property be taxed instead of land? *Regional Science and Urban Economics* 21, 351–370.
- Kanemoto, Y., 1980. *Theories of Urban Externalities*. North-Holland, Amsterdam.
- Kanemoto, Y., 2013a. Second-best cost-benefit analysis in monopolistic competition models of urban agglomeration. *Journal of Urban Economics* 76, 83–92.
- Kanemoto, Y., 2013b. Evaluating benefits of transportation in models of new economic geography. *Economics of Transportation* 2, 53–62.
- Kanemoto, Y., 2013c. Pitfalls in estimating “wider economic benefits” of transportation projects. GRIPS Discussion Paper 13–20.
- Kanemoto, Y., Kitagawa, T., Saito, H., Shioji, E., 2005. Estimating urban agglomeration economies for Japanese metropolitan areas: Is Tokyo too large? In: Okabe, A. (Ed.), *GIS-Based Studies in the Humanities and Social Sciences*. Taylor&Francis, Boca Raton, pp. 229–241.
- Kanemoto, Y., Mera, K., 1985. General equilibrium analysis of the benefits of large transportation improvements. *Regional Science and Urban Economics* 15, 343–363.

- Kanemoto, Y., Ohkawara, T., Suzuki, T., 1996. Agglomeration economies and a test for optimal city sizes in Japan. *Journal of the Japanese and International Economies* 10, 379–398.
- Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71, 1695–1725.
- Mieszkowski, P., Zodrow, G.R., 1989. Taxation and the Tiebout model: The differential effects of head taxes, taxes on land rents, and property taxes. *Journal of Economic Literature* 27, 1098–1146.
- Ogaki, M., Zhang, Q., 2001. Decreasing risk aversion and tests of risk sharing. *Econometrica* 69, 515–526.
- Ottaviano, G.I.P., Tabuchi, T., Thisse, J.-F., 2002. Agglomeration and trade revisited. *International Economic Review* 43, 409–436.
- Papageorgiou, Y.Y., Pines, D., 1999. *An Essay on Urban Economic Theory*. Kluwer Academic Publishers, Boston.
- Romer, P., 1994. New goods, old theory, and the welfare costs of trade restrictions. *Journal of Development Economics* 43, 5–38.
- Schweizer, U., 1983. Efficient exchange with a variable number of consumers. *Econometrica* 51, 575–584.
- Stiglitz, J.E., 1977. The theory of local public goods. In: Feldstein, M.S., Inman, R.P. (Eds.), *The Economics of Public Services*. MacMillan, London, pp. 274–333.
- Tsuneki, A., 1987. The measurement of waste in a public goods economy. *Journal of Public Economics* 33, 73–94.
- Venables, A.J., 2007. Evaluating urban transport improvements: Cost-benefit analysis in the presence of agglomeration and income taxation. *Journal of Transport Economics and Policy* 41, 173–188.
- Vives, X., 1985. On the efficiency of Bertrand and Cournot equilibria with product differentiation. *Journal of Economic Theory* 36, 166–175.
- Zhelobodko, E., Kokovin, S., Parenti, M., Thisse, J.-F., 2012. Monopolistic competition: Beyond the Constant Elasticity of Substitution. *Econometrica* 80, 2765–2784.